ISTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

HIDDEN MARKOV MODELS FOR FAULT CHARACTERIZATION AND CONDITION MONITORING OF INDUCTION MOTORS

M.Sc. Thesis by

H. Erinç KARATOPRAK, B.Sc.

Department : Electrical Engineering

Programme: Electrical Engineering

Supervisor : Prof. Dr. Serhat ŞEKER

JANUARY 2008

ISTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

HIDDEN MARKOV MODELS FOR FAULT CHARACTERIZATION AND CONDITION MONITORING OF INDUCTION MOTORS

M.Sc. Thesis by
H. Erinç KARATOPRAK, B.Sc.
(504061013)

Date of submission: 24 December 2007

Date of defence examination: 30 January 2008

Supervisor (Chairman): Prof. Dr. Serhat ŞEKER

Members of the Examining Committee: Assoc. Prof.Dr. Zehra ÇATALTEPE

Assist. Prof.Dr. Deniz YILDIRIM

JANUARY 2008

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

ASENKRON MOTORLARDA ARIZA ÖZELİĞİ TESBİTİ VE DURUM İZLEME İÇİN SAKLI MARKOV MODELLERİ

YÜKSEK LİSANS TEZİ

Müh. H. Erinç KARATOPRAK
(504061013)

Tezin Enstitüye Verildiği Tarih : 24 Aralık 2007
Tezin Savunulduğu Tarih : 30 Ocak 2008

Tez Danışmanı : Prof. Dr. Serhat ŞEKER
Diğer Jüri Üyeleri Doç. Dr. Zehra ÇATALTEPE
Yrd. Doç. Deniz YILDIRIM

OCAK 2008

ACKNOWLEDGEMENT

I would like to express my deep appreciation and thanks for my family, my advisor Prof. Dr. Serhat Şeker, and Associate Prof. Dr. Zehra Çataltepe, my dear friends Tayfun Şengüler and Merve Acer, my future wife Bilgen Yıldız for their support in both mental and technical aspects.

December 2007

H. Erinç KARATOPRAK

TABLE OF CONTENTS Page

LIST OF TABLES

Page No:

LIST OF FIGURES

Page No:

ABBREVIATIONS

AC : Alternating Current
CWT : Continuous Wavelet Transform
DWT : Discrete Wavelet Transform
EDM : Electrical Discharge Machining
HP : Horse-power
HMM : Hidden Markov Model
MRWA : Multi-resolution Wavelet Analysis
PCA : Principal Component Analysis
WPD : Wavelet Packet Decomposition

HIDDEN MARKOV MODELS FOR FAULT CHARACTERIZATION AND CONDITION MONITORING OF INDUCTION MOTORS

SUMMARY

Induction motors are one of the most common electrical machines with a very wide range of application field from home equipments to nuclear plants. Most of the industrial processes depend on their reliable operation. Predictive maintenance techniques help determine the condition of the online motor in order to predict the condition of the motor to foresee a fault and schedule a maintenance before its occurrence. This offers an advantage of cost saving over routine time-based preventive maintenance because tasks are performed only when necessary. Predictive maintenance aims the goal of predicting the future trend of the equipment's condition which uses principles of statistical process control to determine at what point in future the maintenance will be necessary.

A major component of the predictive maintenance is condition monitoring. Condition monitoring is useful for maintenance scheduling and fault prevention, which is way more cost effective than a failure itself.

Current and vibration analysis are the most commonly used data for the condition monitoring of induction motors, the most common two faults are bearing damage and stator winding faults.

This study has two different applications; fault characterization and condition monitoring.

The first one; fault characterization application uses time-frequency signal analysis methods, Hidden Markov Models and various dimensionality reduction techniques in order to find the frequency band characteristics of a bearing damage on the motor data from an accelerated aging process.

The second study is an application of an online condition monitoring. For this study the accelerated aging motor data is arranged in such a way that a real time condition monitoring system can be tested with it. The condition monitoring system; consisting of a Hidden Markov Model is used with the motor data in order to foresee the bearing damage and stator winding faults both from vibration and current data.

The results of applications have shown that Hidden Markov Models are very accurate and reliable in condition monitoring of induction motors, whereas their usage with dimensionality reduction methods can reveal the physical facts behind the faults.

ASENKRON MOTORLARDA ARIZA ÖZELİĞİ TESBİTİ VE DURUM İZLEME İÇİN SAKLI MARKOV MODELLERİ

ÖZET

Asenkron motorlar günümüzde çok yaygın olarak kullanılan elektrik makinalarıdır. Uygulama alanları evlerdeki buzdolaplarından, nükleer santrallere kadar uzanan çok geniş bir alanı kapsamaktadır. Birçok endüstriyel süreç, asenkron motorların sorunsuz, güvenilir bir şekilde çalışmasına bağlıdır. Öngörülü bakım teknikleri, motorun durumunu izlemek ve oluşabilecek arızaları önceden tespit ederek planlanmış bakımlar yapılmasını sağlar. Bu teknik, sadece arıza riski olduğunda bakım yapılmasını gerektirdiği için, geleneksek periyodik bakıma göre daha ucuz ve etkilidir.

Öngörülü bakımın en önemli parçalarından biri durum izlemedir. Durum izleme, bakım planlanaması ve arızanın önlenmesini sağladığı için oluşan bir arızaya göre çok daha az maliyetlidir.

Asenkron motorlarda en yaygın olarak rulman ve stator sargı arızaları görülür. Akım ve titreşim analizi bu arızalara dair yapılan durum izleme için en çok başvurulan yöntemlerdir.

Bu çalışma arıza özeliği tespiti ve durum izleme olmak üzere iki farklı uygulama içermektedir.

Arıza özeliği tespiti olan ilk uygulamada çok çözünürlüklü dalgacık analizi, Saklı Markov Modeli ve çeşitli boyut azaltma yöntemleri bir arada kullanılarak, rulman arızasına ait fiziksel özelikleri içeren frekans bantları tespit edilmeye çalışılmıştır. Uygulamada hızlandırılmış yaşlandırma ile eskitilmiş motor verileri kullanılmıştır.

İkinci uygulama olan, gerçek zamanlı durum izleme uygulamasında, hızlandırılmış yaşlandırma verileri gerçek zamanlı olarak motordan alınarak, tasarlanan durum izleme sistemine verilecek şekilde düzenlendi. Durum izleme sistemi çok çözünürlüklü dalgacık analizi ve Saklı Markov Modelinden oluşmaktadır. Sistem ayrı ayrı akım ve titreşim verilerini kullanarak rulman ve stator sargı arızalarını öngörülü olarak tespit etmeye çalışmıştır.

Uygulamaların sonuçları, Saklı Markov Modellerinin asenkron motorlarda durum izleme için oldukça başarılı ve tutarlı olduğunu göstermiştir. Aynı zamanda, Saklı Markov Modelleri çeşitli boyut azaltma yöntemleriyle beraber kullanılarak, oluşan arızaların arkasında yatan fiziksel gerçekleri işaret eden frekans bantlarını da tespit edebilmektedir.

1. INTRODUCTION

"I am busy just now again on Electro-Magnetism, and think, I have got hold of a good thing, but can't say; it may be a weed instead of a fish that after all my labor I may at last pull up. I think I know why metals are magnetic when in motion though not (generally) when at rest."Michael Faraday [1].

186 years have passed since the demonstration of the principle of conversion of electrical energy into mechanical energy by electromagnetic means by Michael Faraday. Since then, the electrical machinery have improved a lot. Nowadays, induction motors are one of the most common electrical machines, due to their reliability and relatively low cost. They have a very wide range of application from home equipments like refrigerators, washing machines, air conditioners etc. to intensive care units in hospitals, nuclear reactors, factories etc. They are not only getting popular in the electricity consuming part but also being increasingly used in electricity generation as wind turbines.

It can be said that today's civilization has an increasing dependency on induction machines and their smooth functioning. Reliability of induction machines is a challenging problem of the electrical engineering.

Induction motor is a kind of AC motor, where induction is the power supplier of the rotating device; the rotor, which is the part where electrical energy is converted to mechanical energy.

An induction motor is referred as a rotating transformer because the stator is the primary side and the rotor is the secondary side of the transformer [2].

Induction motors are the popular choice for industry due to their rugged construction, lack of brushes and the ability to easily control the speed due to the new developments in semiconductor devices.

Figure 1.1 and Figure 1.2 show the general structure of an induction motor and its rotor respectively.

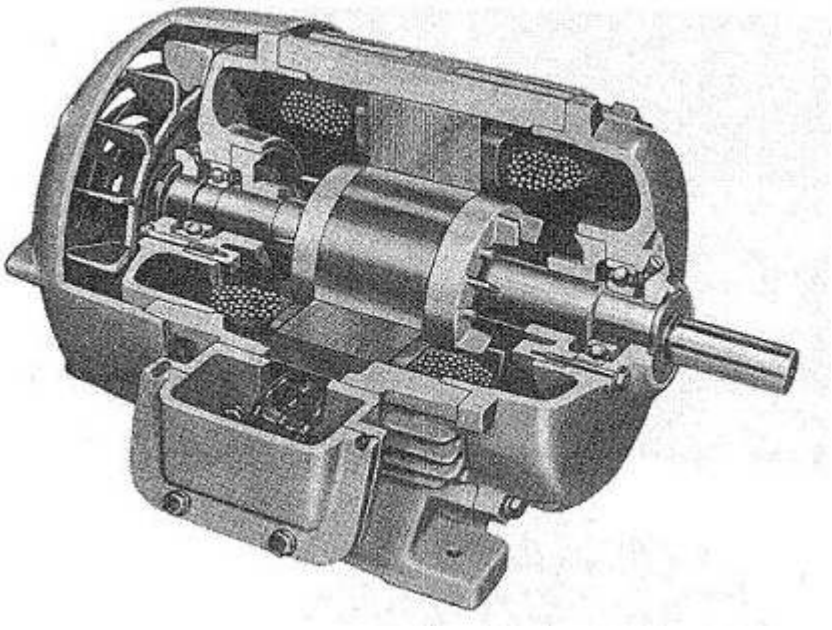

Figure 1.1: Structure of Induction Motor adapted from [2].

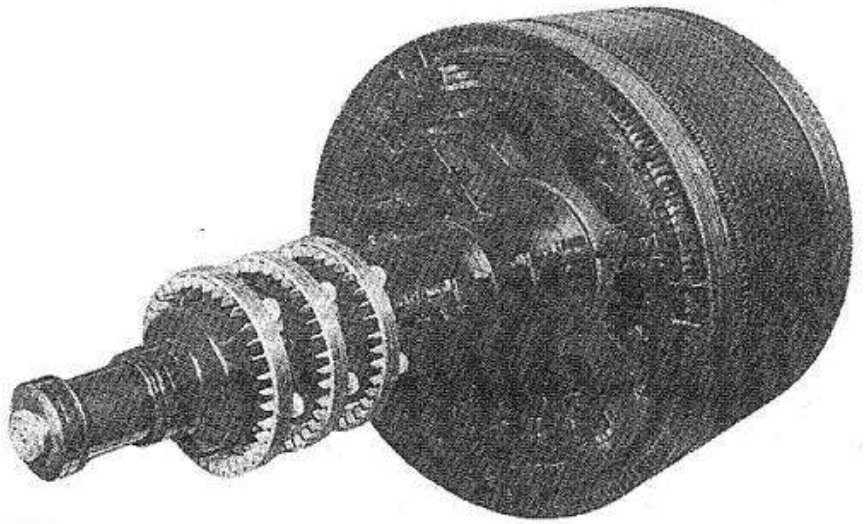

Figure 1.2: Structure of Rotor adapted from [2].

Predictive maintenance techniques help determine the condition of the in-service motor in order to predict when maintenance should be performed. It offers an advantage of cost saving over routine time-based preventive maintenance because tasks are performed only when necessary. Predictive maintenance aims at predicting the future trend of the equipment's condition which uses principles of statistical process control to determine at what point in the future maintenance will be necessary.

A major component of the predictive maintenance is condition monitoring. Condition monitoring is useful for maintenance scheduling and fault prevention, which is way more cost effective than a failure itself.

In induction machines, most of the studies are focused on mechanical faults. The most common failures on induction motors are bearing faults [3]. Amongst these, vibration analysis is the most commonly used one [3].

There have been many approaches for analyzing the vibration signals.

Studies like [4], [5], [6] and [7] have proven that frequency and time-frequency domain analysis techniques give accurate results for fault detection and classification, however they require a supervisor for real-time decision making.

[8], [9] with quick and simple time-domain analysis, and [3] with time-frequency domain analysis, have shown that accurate condition monitoring can be provided with machine learning – pattern recognition techniques like bayes decision algorithm, neural networks and hidden markov models respectively.

This study combines the time-frequency domain analysis technique multi resolution wavelet analysis with the state-of-art tool of speech recognition, namely hidden markov models, and compares various dimensionality reduction techniques to provide a reliable condition monitoring while trying to derive some physical properties using tools of pattern recognition.

This chapter has given a brief introduction to the study. The second chapter provides the mathematical preliminaries like the wavelet, principal component analysis, hidden markov models etc. In the third chapter, an overview of the aging experimental setup and the vibration data acquisition system is given. The fourth chapter is the core of this study, where the application is described in detail. The last chapter gives the conclusions and the possible future works the study would lead to.

2. PATTERN RECOGNITION PRELIMINARIES

2.1 Hidden Markov Models

The output of real-world processes can be considered as signals that can be characterized in various ways, i.e., continuous - discrete, stationary- non stationary, pure-noisy etc. Characterizing those signals with signal models is an important problem of engineering. There are two main methods for creating signal models that characterize the signal that are the deterministic models and the statistical models [10].

Deterministic models use known properties of the signal and the problem is about the estimation-tweaking of the parameter values of the model, whereas statistical models are not directly related with the process itself but tries to characterize the statistical properties of the signal. Hidden Markov Models are in the group of statistical modeling, with the assumption that the signal can be characterized as a parametric random process, and that the parameters of the stochastic process can be estimated precisely[10].

2.1.1 Discrete Markov Processes

Consider a system that is in a state from a set of *N* distinct states: $S_1, S_2, \dots, S_N$, at a time t. Figure 2.1 is an illustration of a system where *N=3*. At discrete time intervals, the system goes through a change of state, based on a set of probabilities related with that state. Let *t*=1,2,... and q_t denote the time instants and the state at time *t* respectively. The methodology is valid for any sequencing like "time" or any other temporal sequence.

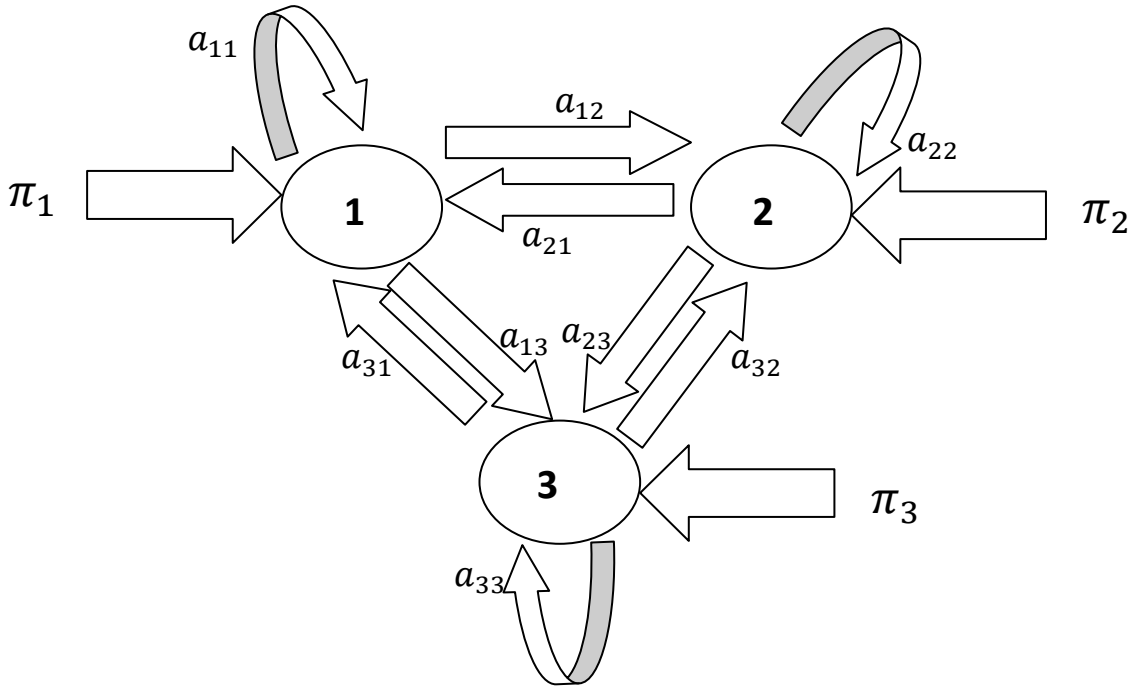

Figure 2.1 : The illustration of a 3 state HMM.

A probabilistic model of the system above would require the knowledge of the state at time *t* (current/instant time) and the previous values, however for a discrete first order Markov chain, the knowledge of the instant and the predecessor state is sufficient, hence according to [10] and [11] the probabilistic description becomes

$$P\left(q_{t+1} = S_j \middle| q_{t-1} = S_i\,, q_{t-2} = S_k\,, \dots\right) = P\left(q_{t+1} = S_j \middle| q_{t-1} = S_i\right) \qquad \textbf{(2.1)}$$

This states that, the future state of the system is only dependent on the current state, the system is independent of the past. The model is further simplified under the assumption of having time invariant transition probabilities [10].

Let a_{ij} show the probability of transition from state *i* to state *j*,

$$a_{ij} = P\left(q_{t+1} = S_j \middle| q_{t-1} = S_i\right), \qquad 1 \le i, j \le N \qquad \textbf{(2.2)}$$

then, the transition coefficients must satisfy the standard stochastic constraints which are as below:

$$a_{ij} \ge 0 \qquad \textbf{(2.3)}$$

$$\sum_{j=1}^{N} a_{ij} = 1 \quad \textbf{(2.4)}$$

In other words, the probability of going from S_i to S_j always has the same probability. The only extension is the initial state. Initial probability π_i is defined as the first state S_i , in a given sequence;

$$\pi_i = \mathrm{P}(\mathrm{q}_1 = \mathrm{S}_\mathrm{i}) \quad \textbf{(2.5)}$$

$$\sum_{i=1}^{N} \pi_i = 1 \quad \textbf{(2.6)}$$

The output of the above process is the set of states at each instant of time, where each state corresponds to an observable event, making it an observable Markov model.

2.1.2 Hidden Markov Models

In a hidden Markov model (HMM), the states are not observable, however an observation is recorded with a probabilistic function of the state [11]. In other words, the states are independent and unobservable variables of the model whereas there are observable variables which are probabilistic functions of those independent variables. The state sequence is estimated using those observation(emission) outputs.

An HMM is specified by the following components:

1) *N*: The number of states in the model

$$S = \{S_1, S_2, \dots, S_N\} \quad \textbf{(2.7)}$$

2) *M*: The number of distinct observation symbols per state(V). The observation symbols correspond to the physical output of the system.

$$V = \{v_1, v_2, \dots, v_M\} \quad \textbf{(2.8)}$$

3) A transition probability matrix, where each a_{ij} represents the probability of transition from S_i to S_j .

$A = [a_{ij}]$ where $a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$

4) A sequence of observation likelihoods, the emission probability that v_m ($m = 1,2, \dots, M$) is observed at S_j .

$B = [b_j(m)]$ where $b_j(m) = P(O_t = v_m | q_t = S_j)$ **(2.9)**

5) Initial state probabilities:

$\boldsymbol{\pi} = [\pi_i]$ where $\pi_i = P(q_1 = S_i)$ **(2.10)**

Since N and M are defined in the other parameters, an HMM can be represented using the below parameter set:

$\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ **(2.11)**

2.1.3 Three Basic Problems of HMMs

The first problem of the HMMs is the evaluation problem, i.e., calculating the probability of an observation sequence, given an HMM model λ.

The second problem is the one in which we attempt to find the hidden part of the model, which means to find the state sequence with the highest probability of generating a given observation sequence.

The third fundamental problem of HMMs can be considered as the training procedure of HMMs. It is to optimize the model parameters so that given a set of observation sequences, the model would maximize the probability of generating that sequences. Mathematically, λ^* is sought after, which maximizes $P(X|\lambda)$. (X is the set of given observation sequences.)

2.1.3.1 Evaluation Problem

Evaluation problem is the calculation of probability of observing the observation sequence *O*, given the state sequence *Q*. If there are *N* different states and the time interval is *T*, then there are N^T possible state sequences, provided that none of the probabilities are zero. This makes the calculations inefficient. Forward-backward procedure is used to calculate $P(O|\lambda)$ efficiently.

Forward-backward procedure divides the observation sequence into two; one being from time 1 to *t*, and the other being from time *t+1* to *T*.

The first section of the observation sequence is defined with the forward variable $\alpha_t(i)$, which is the probability of observing the partial sequence $\{O_1, O_2, \dots, O_t\}$ from time 1 to t, and being in S_i at time t, given the model λ :

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = S_i|\lambda) \quad \textbf{(2.12)}$$

Forward variable α can be calculated recursively;

Step 1: initialization

$$\alpha_1(i) = P(O_1, q_1 = S_i|\lambda) = P(O_1|q_1 = S_i, \lambda)P(q_1 = S_i|\lambda) = \pi_i b_i(O_1) \quad \textbf{(2.13)}$$

Step 2: recursion

$$\alpha_{t+1}(j) = P\big(O_1 \dots O_{t+1}, q_{t+1} = S_j\big|\lambda\big) = \big[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\big] b_j(O_{t+1}) \quad \textbf{(2.14)}$$

$\alpha_t(i)$ explains the first t observations and ends in state S_i . Multiplying this by a_{ij} , to move to the state S_j and since ther are N possible previous states, all such possible previous S_i probabilities need to be summed up. Once the forward variables are calculated, the probability of the observation sequence is given by:

$$P(O|\lambda) = \sum_{i=1}^{N} P(O, q_T = S_i|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \quad \textbf{(2.15)}$$

α_T is the probability of generating the given observation sequence and ending up in state S_i .

2.1.3.2 Finding the State Sequence

The second problem is to find the state sequence $Q = \{q_1 q_2 \dots q_T\}$ with the highest probability of generating a given observation sequence $O = \{O_1 O_2 \dots O_T\}$ with a certain model λ .

To solve this problem, the backward variable of the forward-backward procedure is used.

Backward variable; $\beta_t(i)$ is the probability of being in S_i at time t and observing the partial sequence $O_{t+1} O_{t+2} \dots O_T$.

$$\beta_t(i) = P(O_{t+1} O_{t+2} \dots O_T|q_T = S_i, \lambda) \quad \textbf{(2.16)}$$

The recursive algorithm for calculating β is given below;

Step 1: Initialization (random)

$$\beta_T(i) = 1 \tag{2.17}$$

Step 2: Recursion

$$\beta_t(i) = P(O_{t+1}O_{t+2} \dots O_T | q_T = S_i, \lambda) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \tag{2.18}$$

Let $\gamma_t(i)$ be the probability of being in state S_i at time t, given the observation sequence and λ ;

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(i)\beta_t(i)} \tag{2.19}$$

$\alpha_t(i)$ and $\beta_t(i)$; namely the forward and backward variables divide the sequence between eachother.

Forward variable α is about the sequence from the starting point to time t, S_i ; whereas backward variable β explains the rest from time t to T.

Figure 2.2 shows the forward-backward procedure.

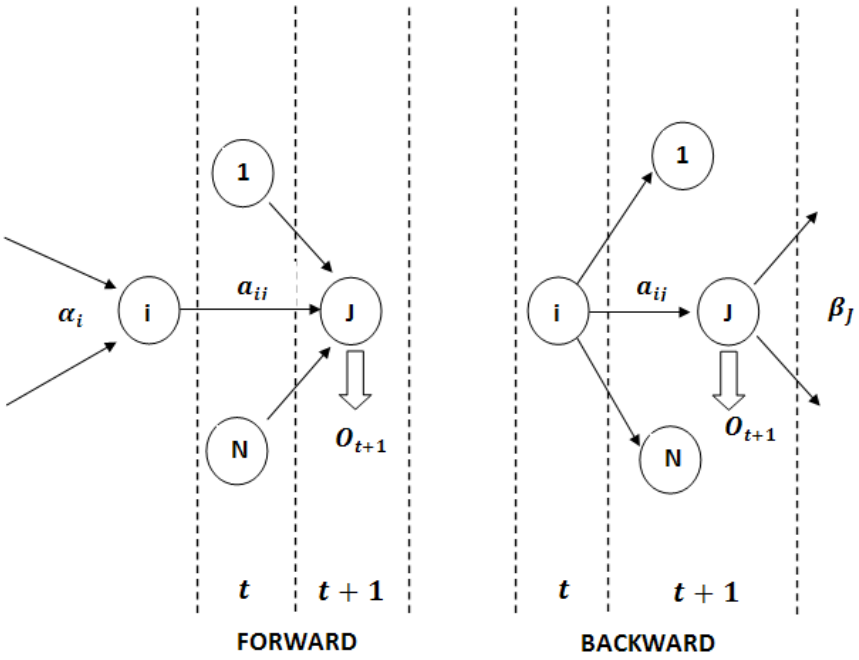

Figure 2.2 : The forward backward procedure

The state sequence at a given step t can be given as :

$$q_t^* = argmax_i \gamma_t(i) \quad \textbf{(2.20)}$$

Which means the state sequence is given by the sequence with the highest probability.

However, this method will bring up the problem of choosing states having a 0 (zero) value transition probability.

A dynamic programming algorithm; Viterbi algorithm, is introduced to overcome this problem [10, 11, 12].

Let $\delta_t(i)$ represent the probability of the highest probability path at time *t* for the first *t* observations ending in S_i , provided that both state and observation sequences are known.

$$\delta_t(i) = max_{q_1 q_2 \dots q_{t-1}} p(q_1 q_2 \dots q_{t-1}, q_t = S_i, O_1 O_2 \dots O_t | \lambda) \quad \textbf{(2.21)}$$

The $\delta_t(i)$ values can be calculated recursively, and the optimal path (Viterbi path) can be backtracked from *T* by choosing the ones with the highest probability at every *t*.

The algorithm is as follows: [10, 13, 14]

Step 1: Initialization

$$\delta_1(i) = \pi_i b_i(O_1) \quad \textbf{(2.22)}$$

$$\psi_1(i) = 0 \quad \textbf{(2.23)}$$

Step 2: Recursion

$$\delta_t(j) = max_i \delta_{t-1}(i) a_{ij} b_j(O_t) \quad \textbf{(2.24)}$$

$$\psi_t(j) = argmax_i \delta_{t-1}(i) a_{ij} \quad \textbf{(2.25)}$$

Step 3: Termination

$$p^* = max_i \delta_T(i) \quad \textbf{(2.26)}$$

$$q_T^* = argmax_i \delta_T(i) \quad \textbf{(2.27)}$$

Step 4: Path (state sequence) backtracking

$$q_T^* = \psi_t(q_{T+1}^*)\,, t = T-1, T-2, \ldots, 1 \quad \textbf{(2.28)}$$

Using the lattice structure of HMMs shown in Figure 2.3 $\psi_t(j)$ keeps the track of the states.

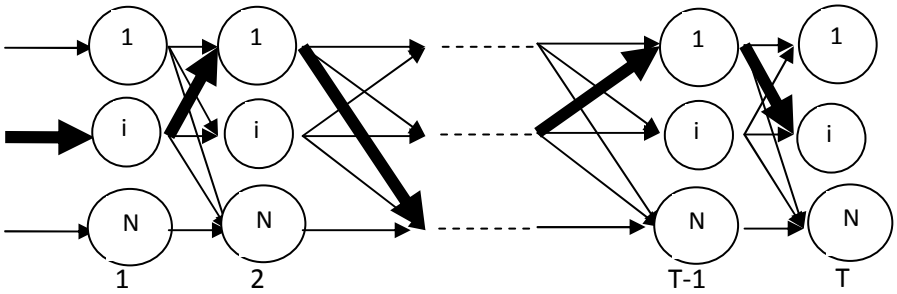

Figure 2.3 : An hmm unfolded in time as a lattice showing possible paths. The path with thick lines is the state sequence path that made the given emission sequence.

2.1.3.3 Learning an Hmm from data

The third problem is about finding a method to set the model parameters (A, B, π) to maximize the observation probability, given the model. However there is not an analytical solution for the model to maximize the probability of the observation sequence. Baum-Welch algorithm can be used to find a local maximum [15, 16].

Maximum likelihood is used to calculate λ^* which maximizes the likelihood of the sample of training sequences $X = \{O^k\}_{k=1}^K$.

Let $\xi_t(i,j)$ be the probability of being in S_i at time *t* and in S_j at time *t+1* given O and λ .

$$\xi(i,j) = P\left(q_t = S_i, q_{t+1} = S_j \middle| O, \lambda\right) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_k \sum_l \alpha_t(k)a_{kl}b_l(O_{t+1})\beta_{t+1}(l)} \quad \textbf{(2.29)}$$

The calculation of the probability of being in state S_i at time *t* can be given by:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i,j) \quad \textbf{(2.30)}$$

Baum-Welch algorithm is an iterative expectation maximization procedure, that is composed of expectation and maximization steps namely E-step and M-step.

At each iteration, first in the E-step, $\xi_t(i,j)$ and $\gamma_t(i)$ are computed based on that instant $\lambda = (A, B, \pi)$ model.

Secondly in the M-step, λ is recalculated given $\xi_t(i,j)$ and $\gamma_t(i)$.

E-M steps are repeated until the convergence where $P(O|\lambda)$ stops decreasing.

Let z_i^t and z_{ij}^t be the indicator variable;

$$z_i^t = \begin{cases} 1 \; if \; q_t = S_i \\ 0 \quad otherwise \end{cases} \tag{2.31}$$

$$z_{ij}^t = \begin{cases} 1 \; if \; q_{t+1} = S_j \\ 0 \quad otherwise \end{cases} \tag{2.32}$$

They are estimated in the E-step:

$$E[z_i^t] = \gamma_t(i) \tag{2.33}$$

$$E\left[z_{ij}^t\right] = \xi_t(i,j) \tag{2.34}$$

Using these estimations, parameters are calculated in the M-step.

The probability of transition from S_i to S_j can be calculated by dividing the expected number of transitions from S_i to S_j $(\sum_t \xi_t(i,j))$ to the total number of transitions from S_i $(\sum_t \gamma_t(i))$.

$$\widehat{a_{ij}} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{2.35}$$

The probability of observing v_m in S_j is the expected number of times v_m is observed when the system is in S_j over the total number of times the system is in S_j [11].

$$\widehat{b_j}(m) = \frac{\sum_{t=1}^{T} \gamma_t(j) 1(O_t = v_m)}{\sum_{t=1}^{T} \gamma_t(j)} \tag{2.36}$$

Multiple observation sequences

$$X = \{O^k\}_{k=1}^K \tag{2.37}$$

which is assumed to be independent

$$P(X|\lambda) = \prod_{k=1}^{K} P(O^k|\lambda) \quad \textbf{(2.38)}$$

The parameters are now averages over all observations in all sequences.

$$\widehat{a_{ij}} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k-1}\xi_t^k(i,j)}{\sum_{k=1}^{K}\sum_{t=1}^{T_k-1}\gamma_t^k(i)} \quad \textbf{(2.39)}$$

$$\widehat{b_j}(m) = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k-1}\gamma_t^k(j)1(O_t^k=v_m)}{\sum_{k=1}^{K}\sum_{t=1}^{T_k-1}\gamma_t^k(j)} \quad \textbf{(2.40)}$$

$$\widehat{\pi_i} = \frac{\sum_{k=1}^{K}\gamma_1^k(i)}{K} \quad \textbf{(2.41)}$$

2.1.4 Types of HMMs

The HMM should be tuned to be appropriate for the dataset. Some applications require a fully connected (ergodic) HMM while in some applications only certain transitions exist [17].

In applications like speech recognition and aging assessment, left-to-right HMMs as shown in Figure 2.4 are used, where the states have time order. As the time increases, the state index either does not change or increases, which means there are no possible backward transitions [11, 18].

Mathematically:

$$a_{ij} = 0 \ \forall \ j < i \quad \textbf{(2.42)}$$

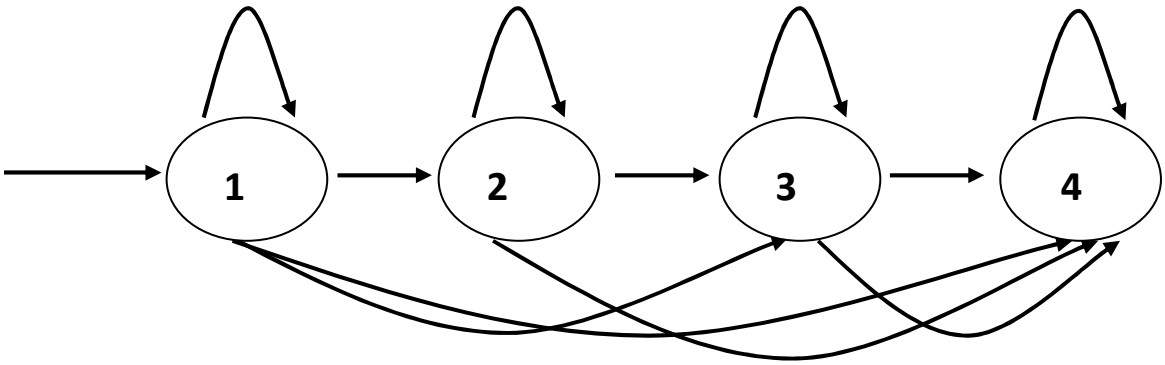

Figure 2.4 : A left to right HMM

2.2 Dimensionality Reduction and Principal Component Analysis

"If Edison had a needle to find in a haystack, he would proceed at once with the diligence of the bee to examine straw after straw until he found the object of his search. I was a sorry witness of such doings, knowing that a little theory and calculation would have saved him ninety per cent of his labor." Nikolai Tesla [19].

2.2.1 Dimensionality Reduction

Dimensionality reduction is the process of reducing the number of inputs(random variables). It is useful because:

- Complexity comes from the input dimension in most of the learning algorithms.
- Reduced memory and computation : cost effective
- Removing a feature means to save the cost of extracting that feature (e.g. removal of a sensor)
- More robust and stable simplified models.
- It leads to knowledge extraction [11].
- Helps to understand and analyze the data better by allowing a visual plotting of the data in some cases.

Dimensionality reduction can be divided into two categories; feature selection and feature extraction[20-21]. Feature selection methods try to find a subset of the features that explain the data well, thus the rest of the features can be removed[20-21].

Feature extraction methods produce a mapping of the feature space into a new one with less dimensions [20-21].

Principal Component Analysis is one of the most common feature extraction method in which the original feature space is projected to a new space with fewer dimensions [11,20,21].

2.2.2 Principal Component Analysis

Principal component analysis is a projection method which has roots in statistics and applied linear algebra. It is a simple, non-parametric method for extracting information from complex data sets [22].

The focus of PCA is to find a mapping from the inputs in the original d-dimensional space to a new (k<d)-dimensional space, with as little as possible loss of information [11].

It involves the calculation of the eigenvalue decomposition of a dataset to achieve the goal of computing the most significant basis to remove the noise and reveal the hidden structure [22]. The results are discussed in terms of component scores and loadings.

Mathematically, PCA is an orthogonal linear transformation that transforms the data to a new coordinate system, where the greatest variance by any projection of the data comes to lie on the first coordinate called the first principal component, the second greatest variance on the second coordinate and so on. PCA can be used for dimensionality reduction in a data set, by keeping the characteristics which contribute most to its variance while removing the rest. However, based on the application, the highest variance characteristics needn't always include the "most important" aspects of the data [23].

Let X be an MxN data matrix, where *M* is the number of experiences and *N* is the number of features(dimensions). *Y* is an *LxN* matrix, where every column represents a projection from matrix *X* onto the basis vectors contained in the columns of matrix *W*.

Then the PCA transformation of *X* is;

$$Y = W^T X \quad \textbf{(3.1)}$$

The algorithm for dimensionality reduction using PCA is [23]:

1) Subtract the mean from the collected data.

2) Calculate the covariance matrix.

3) Calculate the eigenvectors and eigenvalues of the covariance matrix.

4) Choose the components and form a feature vector with the chosen eigenvectors(components). Here, the eigenvector with the highest eigenvalue is the principle component. Components of lesser significance can be ignored, which is dimensionality reduction.

5) Derive the new data set;

Newdataset=RowFeatureVector x$Newdataset = RowFeatureVector\ x\ (data)^T$

The "Newdataset" matrix will give the data with reduced features (dimensions).

2.3 K-means Clustering

Clustering is the classification of objects into different groups i.e., the partitioning of a data into subsets called clusters, so that the data in each subset share some common feature, which is often a distance measure [24].

The pattern recognition techniques can be grouped in two, namely parametric and non-parametric methods. In parametric methods, the sample is assumed to come from a known distribution.

However in non-parametric methods, such an assumption is not acceptable [11]. In a non-parametric case where also there are not known classes to put the data into (unsupervised) , clustering is used to create optimal subsets by the data in hand.

K-means clustering is a clustering algorithm for partitioning *N* data points into *K* disjoint subsets S_j containing N_j data points so as to minimize the sum-of-squares criterion:

$$J = \sum_{j=1}^{K} \sum_{n \in S_j} |X_n - \mu_j|^2 \qquad \textbf{(4.1)}$$

where x_n is a vector denoting the nth data point and μ_j is the geometric centroid of the data points in S_j [25].

The algorithm has a simple re-estimation method. First the data points are randomly assigned to the *K* sets. In the first step, the centroids of each set are calculated. In the second step, each point is assigned to sets with the closest centroid. These two steps are repeated until there is not a change in the assignments [Url-1].

3. SIGNAL ANALYSIS

Signal analysis is the analysis, interpretation and manipulation of signals. Storage, reconstruction, separation of information from noise, compressing and feature extraction are the main branches of signal processing [29].

3.1 Time Domain Analysis

Time domain analysis focuses on the temporal characteristics of a signal over a definite time period. Tools of time domain analysis are integration, differentiation, finding the correlation of two signals and using differential equations to describe a physical process [29].

One of the most important methods is the convolution which is the integrated product of two functions that have been time-shifted. The impulse response of the system is used with convolution to describe a physical system [29].

The continuous convolution of a signal can be given by;

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \quad \textbf{(3.1)}$$

The discrete convolution can be given by;

$$y[n] = x[n] * h[n] = \sum_{-\infty}^{\infty} x[k]h[n-k] \quad \textbf{(3.2)}$$

3.2 Orthogonal and Integral Transforms

Orthogonal transforms depend on a set of orthogonal vectors or a set of orthogonal functions called basis [30].

These transforms provides the study of components that are mapped onto the basis instead of studying the signal sequence or function. Briefly; by inner product, signals are mapped onto the basis functions that are orthogonal [29].

3.2.1 Fourier Analysis

Frequency analysis is based on the Fourier series, where any arbitrary periodic function can be expressed as an infinite series of weighted sinusoidal functions [31].

The basis functions are $sin(kx)$ and $cos(kx)$ which are orthogonal. The weight coefficients (D_k) are given by [31];

$$D_k = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_t(t)e^{-jk\omega t}\,dt \quad \textbf{(3.3)}$$

where k is the frequency and T is the period.

The Fourier series of a signal is;

$$x_T(t) = \sum_{k=-\infty}^{\infty} D_k e^{jk\omega t} \quad \textbf{(3.4)}$$

3.2.1.1 Fourier Transform

The above only applies for periodic signals. In order to represent non-periodic signals, an integral transform - Fourier Transform is used [31]:

$$X(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt \quad \textbf{(3.5)}$$

where the complex function $X(\omega)$ is the spectral density, a continuous distribution of the sinusoid.

The inverse Fourier transform is;

$$x(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}d\omega \quad \textbf{(3.6)}$$

Fourier transform for a discrete signal can be given by;

$$G\left(e^{j\omega}\right) = \sum_{n=-\infty}^{\infty} g(n)e^{-j\omega n} \quad \textbf{(3.7)}$$

The inverse Fourier transform of a discrete time sequence is;

$$g(n) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} G(e^{j\omega})e^{j\omega t}d\omega \quad \textbf{(3.8)}$$

3.2.1.2 Discrete Fourier Transform (DFT)

Discrete Fourier Transform (DFT) is implemented by sampling the discrete points on the spectrum. Basis functions are discrete sinusoids. DFT treats $x[n]$ as a periodic function with the period *N*.

The analysis equation can be given by [29];

$$D[k] = \frac{1}{N}\sum_{k=0}^{N-1} x[n]e^{-jk\Omega n} \quad \textbf{(3.9)}$$

where Ω is the discrete angular frequency and k is the index of that Ω collateral to $n\omega$.

Inverse DFT is ;

$$x[n] = \sum_{k=0}^{N-1} D[k]e^{jk\Omega n} \quad \textbf{(3.10)}$$

There are fast and effective algorithms called Fast Fourier Transforms (FFT) to compute DFT and its inverse. Cooley-Tukey, Prime-factor, Bruun's, Rader's and Bluestein's FFT algorithms are among the most common ones. There is also another FFT called Damn Fast Fourier Transform(DFFT) which is very fast but inaccurate [32].

3.2.2 The Laplace and Z-transform

Laplace transform gives information about the stability of a system and the initial conditions.

The complex Laplace transform is defined as follows [29];

$$X(s) = \int_{-\infty}^{\infty} x(t)e^{-st}dt \quad \textbf{(3.11)}$$

The inverse Laplace transform is;

$$x(t) = \frac{1}{2\pi}\int_{c+j\infty}^{c-j\infty} X(s)e^{st}ds \quad \textbf{(3.12)}$$

When analyzing a discrete system, the Z transform is used. It is;

$$H(z) = \sum_{n=-\infty}^{\infty} h[n]z^{-n} \quad \textbf{(3.13)}$$

The inverse z-transform is given by;

$$x[n] = \frac{1}{2\pi j} \oint X(z) z^{n-1} dz \quad \textbf{(3.14)}$$

3.3 Wavelet Transform

Wavelets are similar to Fourier transforms, but instead of the spectral density of a signal, the focus is on the projection of the signal on to functions called wavelets. It divides the given function into different frequency components and studies each component with a resolution that matches its scale.

Wavelet transforms are better than Fourier transforms in terms of functions with discontinuities, sharp peaks,accurate deconstruction and reconstruction of finite, non-periodic and/or non-stationary signals.

3.3.1 Wavelet Theory

All wavelet transforms are time-frequency representation and they are related to harmonics analysis. Almost all of the wavelet transforms use filter banks, which are either finite impulse response (FIR) or infinite impulse response (IIR) filters [33].

There are three groups of wavelet transforms, that are continuous, discrete and multi resolution.

Continuous wavelet transforms are governed by the Heisenberg's uncertainty principle [33].

3.3.1.2 Continuous Wavelet Transform

A signal with finite energy is projected on a continuous group of frequency bands (or similar subspaces of the function space $L^2(\mathbb{R})$), namely on every frequency band on $[f, 2f]$ for all $f > 0$.

The original signal is reconstructable by an appropriate integration over the whole frequency components [34].

The frequency bands or subspaces are scaled versions of a subspace with scale l, which is generated by the shifts of the generating function $\psi \in L^2(\mathbb{R})$ namely the mother wavelet [33].

$$\psi(t) = 2sinc(2t) - sinc(t) = \frac{\sin(2\pi t) - \sin(\pi t)}{\pi t} \quad \textbf{(3.15)}$$

where $sinc$ is the sinus cardinalis function

$$sinc(x) = \frac{\sin(\pi x)}{\pi x} \quad \textbf{(3.16)}$$

The frequency band $[\frac{1}{a}, \frac{2}{a}]$ or the subspace of scale a is generated by the functions called child wavelets.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\, \psi(\frac{t-b}{a}) \quad \textbf{(3.17)}$$

where $a > 0$ is the scale and $b \in \mathbb{R}$ is the shift.

(a, b) is a point in the halfplane $\mathbb{R}_+ x \mathbb{R}$.

The projection of a function x onto the subspace of scale a is;

$$X_a(t) = \int_{\mathbb{R}} WT_\psi\{x\}(a,b).\psi_{a,b}(t)db \quad \textbf{(3.18)}$$

where the wavelet coefficients are [34];

$$WT_\psi\{x\}(a,b) = < x, \psi_{a,b} > = \int_{\mathbb{R}} X(t)\overline{\psi_{a,b}(t)}dt \quad \textbf{(3.19)}$$

To analyze the signal; wavelet coefficients are assembled into a scaleogram of the signal. The wavelet coefficients are similar to the spectral density but it is a surface. Identifying signal characteristics at different points a,b ; a good signal analysis can be made.

Let the energy function E be defined as;

$$E(a,b) = |T(a,b)|^2 \quad \textbf{(3.20)}$$

Any portion of the signal can be zoomed using different scales(a) and time window can be changed by altering b(dilation). Below is an example of a signal analysis by continuous wavelet transform.

Let the signal be $y(t) = \cos(2 * 270 * \pi * t) + \sin(2 * 320 * \pi * t)$ with a normally distributed noise. The plot of the signal, its normalized PSD and CWT can be seen in the below figures Figure 3.1 , Figure 3.2 and Figure 3.3 respectively.

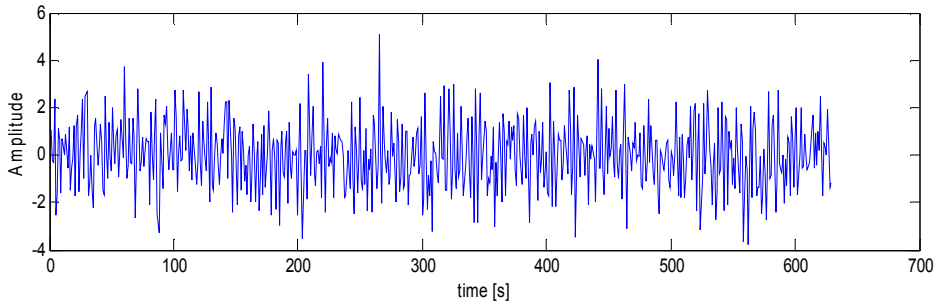

Figure 3.1 : The plot of the given function

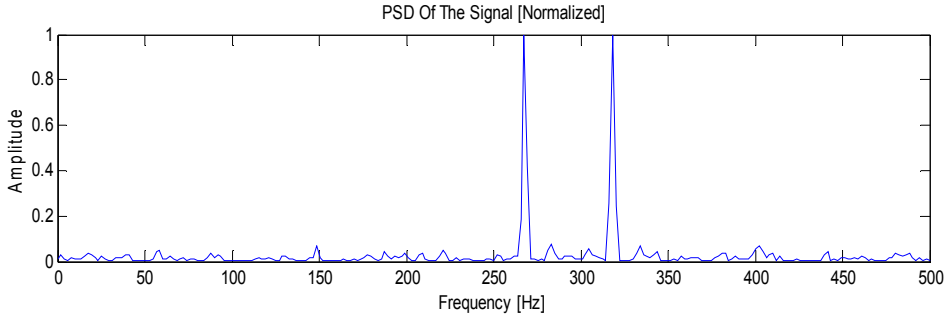

Figure 3.2 : The PSD of the given function

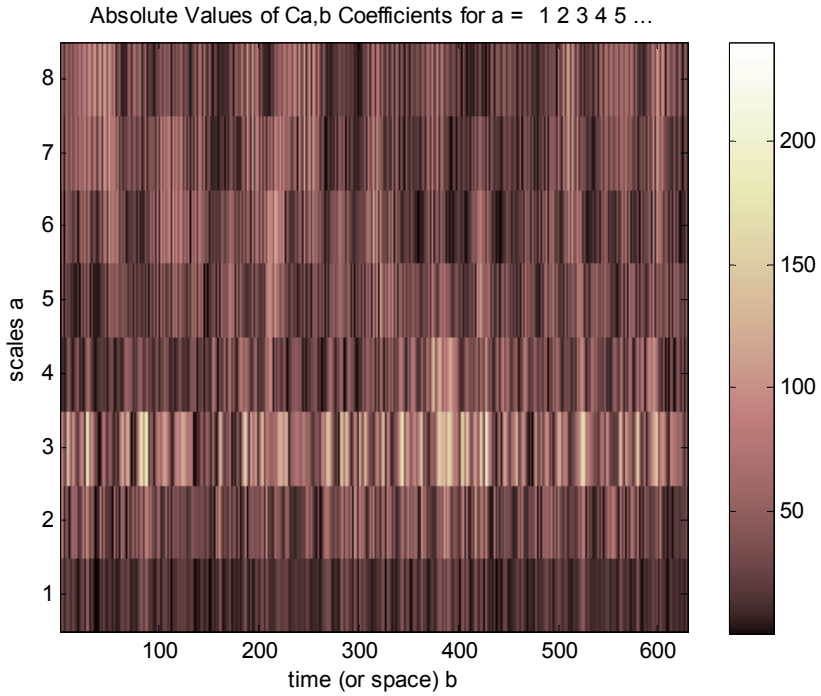

Figure 3.3 : The CWT of the given function

3.3.1.3 Discrete Wavelet Transform

CWT is a good method to analyze the signal, however when it comes to the reconstruction of the signal it might not be that good.

It is computationally impossible to analyze a signal using all of the wavelet coefficients. However it is sufficient to pick a discrete subset of the upper halfplane $\mathbb{R}_+x\mathbb{R}$ in order to reconstruct the signal using those wavelet coefficients.

Let $a > 1$ and $b > 0$, then the discrete subset includes the points $(a^m, na^m b)$ where $m, n \in \mathbb{Z}$. The corresponding baby wavelets can be given as [34];

$$\psi_{m,n}(t) = a^{-\frac{m}{2}}\psi(a^{-m}t - nb) \quad \textbf{(3.21)}$$

And the sufficient condition for the reconstruction of a finite energy signal x with the below formula[35] :

$$x(t) = \sum_{m\in\mathbb{Z}}\sum_{n\in\mathbb{Z}} < x, \psi_{m,n} >. \psi_{m,n}(t) \quad \textbf{(3.22)}$$

is that the functions $\{\psi_{m,n}: m, n \in \mathbb{Z}\}$ form a tight frame of $L^2(\mathbb{R})$ [35].

3.3.1.4 Multi-resolution Wavelet Analysis

In discrete wavelet transform, there are only a finite number of wavelet coefficients however each coefficient requires an integral evaluation. In order to overcome this calculations, an auxiliary function, the father wavelet $\Phi \in L^2(\mathbb{R})$ is used, with the restriction of a being an integer [33,34].

Let $a = 2$ and $b = 1$, which are the popular father and mother wavelets Daubechies 4 tap.

The below subspaces are constructed with the mother and father wavelets

$$V_m = span(\Phi_{m,n}: n \in \mathbb{Z}) \text{ where } \Phi_{m,n}(t) = 2^{-\frac{m}{2}}\Phi(2^{-m}t - n) \quad \textbf{(3.23)}$$

and

$$W_m = span(\psi_{m,n}: n \in \mathbb{Z}) \text{ where } \psi_{m,n}(t) = 2^{-\frac{m}{2}}\psi(2^{-m}t - n) \quad \textbf{(3.24)}$$

Thus, it is required that the sequence

$$\{0\} \subset \cdots \subset V_1 \subset V_0 \subset V_{-1} \subset \cdots \subset L^2(\mathbb{R}) \quad \textbf{(3.25)}$$

Forms a multi resolution analysis of $L^2(\mathbb{R})$, and that the subspaces $\ldots, W_1, W_0, W_{-1}, \ldots$ are the orthogonal "differences" of the above sequence i.e. , W_m and V_m are the orthogonal complements in the subspace V_{m-1} .

As a result of the above relations, there exists two sequences namely $h = \{h_n\}_{n\in\mathbb{Z}}$ and $g = \{g_n\}_{n\in\mathbb{Z}}$ with the following properties [36];

$$h_n = < \Phi_{0,0} , \Phi_{1,n} > \text{ and } \Phi(t) = \sqrt{2} \sum_{n\in\mathbb{Z}} h_n \Phi(2t - n) \quad \textbf{(3.26)}$$

and

$$g_n = < \psi_{0,0} , \Phi_{1,n} > \text{ and } \psi(t) = \sqrt{2} \sum_{n\in\mathbb{Z}} g_n \Phi(2t - n) \quad \textbf{(3.27)}$$

Below Figure 3.4 shows the multi resolution wavelet analysis (MRWA) of the same function used in the above example

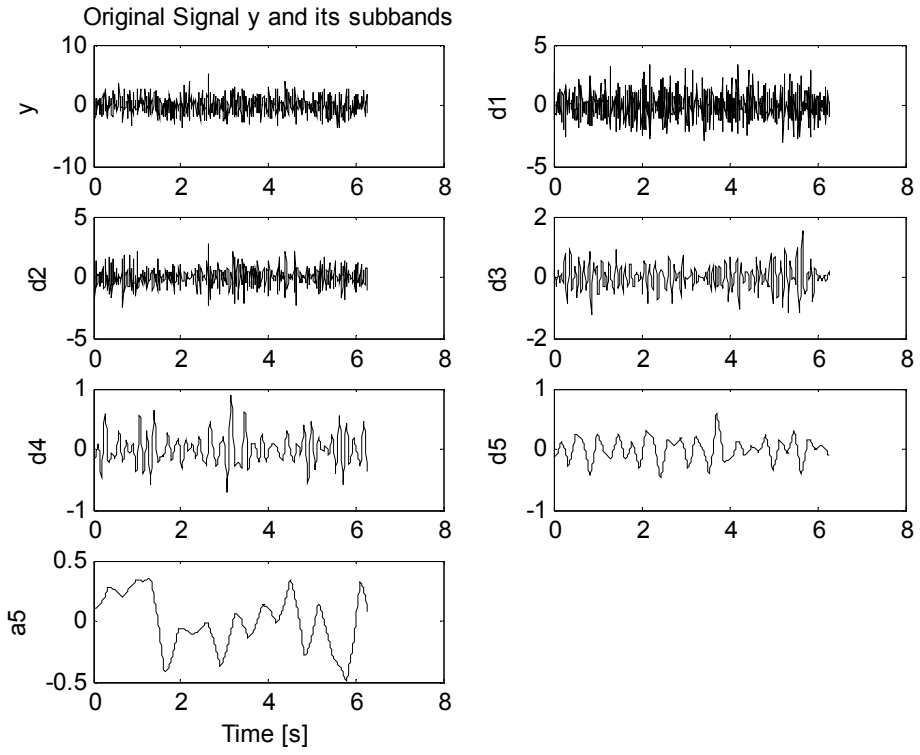

Figure 3.4 : MRWA of the given signal where y is the original signal, d's and a5 are the sub bands, being the details and the approximation respectively.

3.3.2 Mother Wavelet

Mother wavelets are chosen to be continuously differentiable for practical and efficient calculations. However due to analytical requirements, the wavelet functions are chosen from a subspace of the space $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, where every function is absolutely and square integrable [36];

$$\int_{-\infty}^{\infty} \psi|t|dt < \infty \quad \textbf{(3.28)}$$

and

$$\int_{-\infty}^{\infty} \psi|t|^2 dt < \infty \quad \textbf{(3.29)}$$

This space provides the calculation of the conditions of zero mean and square norm one;

Condition for zero mean:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \quad \textbf{(3.30)}$$

Condition for square norm one:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1 \quad \textbf{(3.31)}$$

In CWT, for ψ to be a wavelet transform, the mother wavelet must satisfy the admissibility criterion in order to get a stably invertible transform [33,34].

3.3.2.1 The Admissibility condition for CWT

The CWT of a function f is:

$$\gamma(b,a) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \overline{\psi(\frac{t-b}{a})} dt \quad \textbf{(3.32)}$$

where b is translation, a is scale, ψ is the mother wavelet and $\bar{\psi}$ is the complex conjugate of ψ.

The reconstruction of f can be given by;

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \gamma(b,a) \frac{1}{\sqrt{|a|}} \psi(\frac{t-b}{a}) db \frac{da}{a^2} \quad \textbf{(3.33)}$$

where

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi \quad \textbf{(3.34)}$$

is called the admissibility constant and $\hat{\psi}$ is the fourier transform of ψ .

In order to have an inverse transform the admissibility condition must be satisfied;

$0 < C_\psi < +\infty$ which implies that $\hat{\psi}(0) = 0$, and it means that a wavelet must integrate to zero [33,34].

In DWT, it is required that the wavelet series is a representation of the identity in the space $L^2(\mathbb{R})$.

Most of the DWT constructions use the multi resolution analysis, which defines the wavelet by a scaling function and that function itself is solution to a functional equation [34].

In most situations, restricting ψ to be a continuous function with a higher number M of vanishing moments is useful: for $m < M$

$$\int_{-\infty}^{\infty} t^m \psi(t) dt = 0 \quad \textbf{(3.35)}$$

Below in Figure 3.5 are three popular mother wavelets, Morlet, Mexican Hat and Meyer.

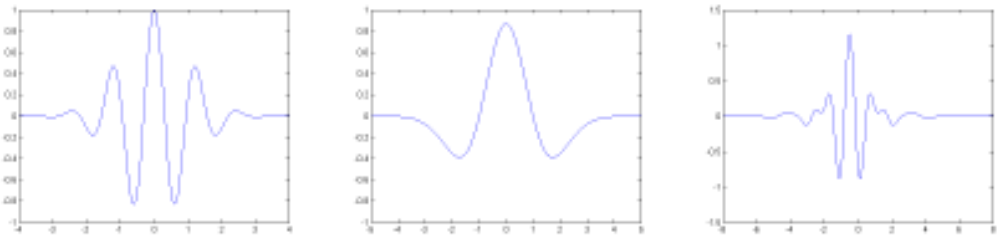

Figure 3.5 : Examples of mother wavelets; Morlet, Mexican Hat and Meyer respectively.

The mother wavelet is scaled (dilated) by a factor of a , and translated (shifted) by a factor of b.

3.4 Shannon Entropy

Shannon Entropy is a measure of uncertainty associated with a random variable. It is a way of quantifying the information contained in a piece of data [26].

The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty [27].

The entropy of a discrete random variable X, that can take on possible values $\{x_1 \dots x_n\}$ is;

$$H(x) = E\big(I(X)\big) = \sum_{i=1}^{n} p(x_i) log_2 \left(\frac{1}{p(x_i)}\right) = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i) \quad \textbf{(3.36)}$$

where I(x) is the information content, which is itself a random variable

$p(x_i) = \Pr(X = X_i)$ is the probability mass function of X.

Let $p_i = \Pr(X = x_i)$ and $H_n(p_1, \dots p_n) = H(X)$

Shannon entropy has the following properties [28]:

- Continuity: The measure must be continuous; changing one of the probability values by a very small amount only changes the entropy by a small amount.
- Symmetry: The measure is not affected by the order of the outcomes x_i .

  $$H_n(p_1, p_2, \dots) = H_n(p_2, p_1, \dots) \quad \textbf{(3.37)}$$

- Maximum: The entropy is maximum when all the outcomes are equally likely. This means that the uncertainty is highest when all possible events have the same probability.
- Additivity: The amount of entropy must be the same regardless of the division of the process into parts. This means that the entropy of a system can be calculated from its subsystems.

4. EXPERIMENT AND APPLICATION

4.1 Experimental Setup and Measurement System

The main reason of the bearing damage is the shaft currents. Shaft currents which are produced by power electronics motor drives are the main reason of the bearing damage [37].

4.1.1 Electrical Discharge Machining

In order to understand the physics of the bearing damage, one must have knowledge on electrical discharge machining. Actually EDM is an accurate manufacturing process for creating complex or simple shapes and geometries within parts and assemblies [38].

EDM is also referred to as spark machining or spark encoding. Briefly, it is a method of removing material by a series of rapidly occurring electric arcing discharges between an electrode namely the cutting tool and the work piece, in the presence of an energetic electric field.

In an induction machine, the rotor is supported by bearings with a dielectric grease film which creates a layer between the balls and the race. During high speed operation, the grease film has a uniform distribution between the rolling elements, the inner and outer races which creates dielectric layers between them. When the rotor voltage increases, at some point, the voltage causes an arc, a discharge current that flows through the bearing elements. This occurrence is very similar with EDM and damages the bearing surfaces.

Especially at low speed operation, the grease film cannot create a good insulation and the grease layer is minimized. In this situation, balls and the race usually have full contact which prevents a voltage increase. Instead of an arcing current, a normal conducting current flows which does not cause any bearing damage.

4.1.2 Experimental Setup for Aging

The test setup design at [39] was used for the simulation of the electrical discharge from the shaft to the bearing. Figure 4.1 shows this setup.

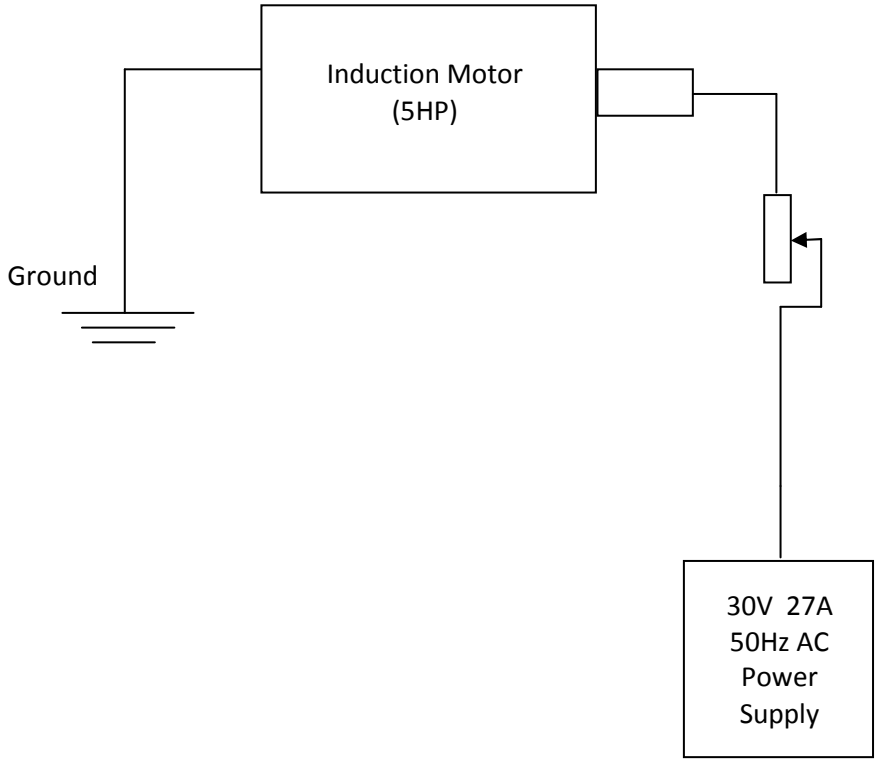

Figure 4.1 : The aging test setup

The experiment was repeated in cycles. The first cycle i.e. cycle 0 is the healthiest cycle (a brand new induction motor) where the last cycle i.e. cycle 7 is the most faulty one (the one just before the motor stopped operating). For every aging cycle (EDM aging cycle), the motor was run at no load for 30 minutes with the external shaft current of 27 Amperes at 30 Volts AC.

In order to accelerate the aging, thermal aging was applied after every EDM cycle [40].

After each cycle, motor was put on a performance test platform, where motor currents and voltages, rotor speed, torque and vibration measurements were recorded with a data acquisition system of 12 kHz. This acquisition system can be seen in Figure 4.2.

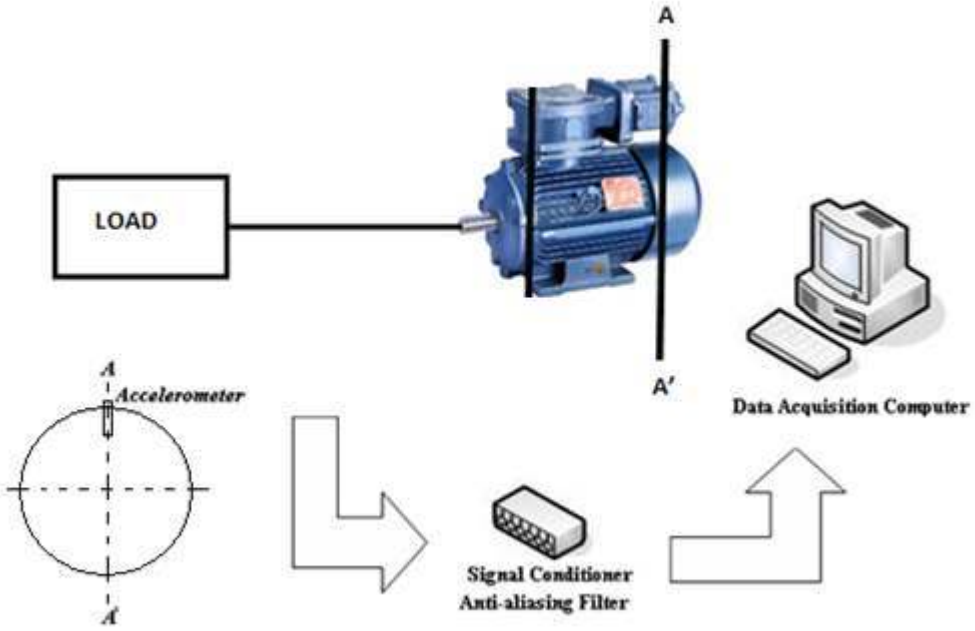

Figure 4.2 : The performance test platform with the data acquisition system

4.2. Application

There are two different applications in this study.

The first application is a laboratory test in which tools of pattern recognition are used to examine a developing fault, and the fault characteristics are extracted, i.e., the exact frequency bands that are related with the given fault are found. One can ask whether the frequency band detection can be made only with signal analysis techniques, or is it necessary to use pattern recognition methods. In some applications the exact frequency bands can be detected visually, however in some applications it is required to use more precise methods.

Briefly the vibration signal of an induction motor is preprocessed and taught to an HMM model from cycle 0 to cycle 7 in which a bearing damage occurs. The success rate of HMM is checked by using different dimensionality reduction techniques in order to come to a conclusion of a certain frequency band where the success to dimension ratio is closest to 1.

The second application is a condition monitoring technique which is applicable to industrial processes in order to provide predictive maintenance. The technique foresees the faults - the rotor and stator faults (which is the deformation of the insulation material of the stator windings) - before they occur. The foreseeing can be

made either by using vibration signals or current signals. The normal condition of the motor is taught to an HMM and with a threshold, the fault is foreseen during operation.

4.2.1 Fault Characterization

In this application, a motor is exposed to the explained EDM process in order to examine a developing bearing damage in 8 cycles, where the last cycle is the point just before the motor can no longer operate. The psd and time-domain plots of the 6 second samples at no load for cycles 0(healthy) and 7(aged) can be seen in Figure 4.3.

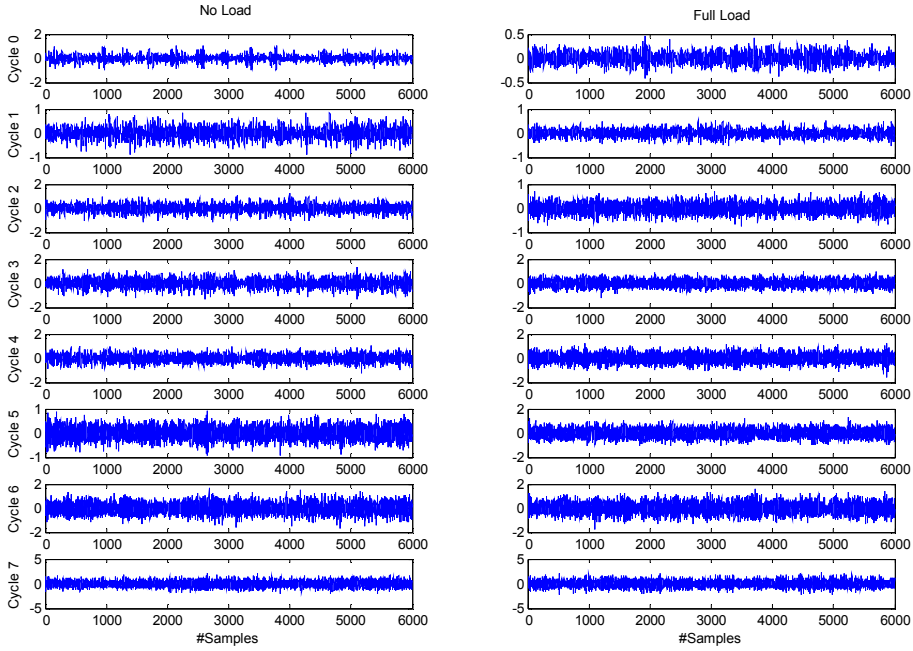

Figure 4.3 : The time domain plots and the psd's of 8 cycles; where 6000 samples make 0.5 seconds with the 12 kHz sampling frequency.

The load of the motor takes on the values of 0%(no-load) , 25% , 50% , 75% , 100%(full-load) of the nominal load. A total vibration measurement time of 10 seconds, 8 seconds for training and 2 seconds for validation is used. Every cycle has 5 different load conditions as described above.

A three level multi resolution wavelet analysis (MRWA) is applied to the signal as seen in Figure 4.4 and the Shannon entropy of each level is taken as a feature forming the 7 dimensional feature vectors.

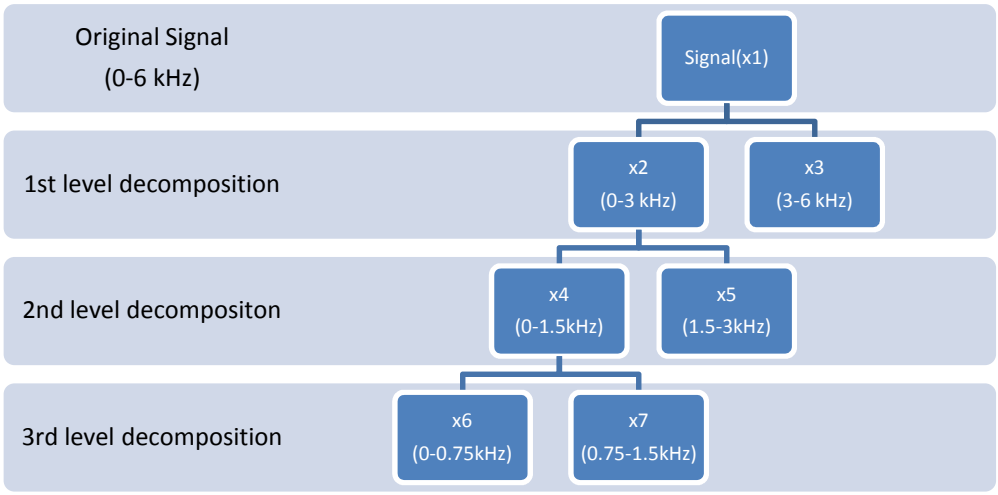

Figure 4.4 : 3 level MRWA decomposition of the signal forming the feature vectors

Figure 4.5 shows the MRWA of 2 second samples of the vibration signal from the cycles 0(healthy) and 7(aged) at no load.

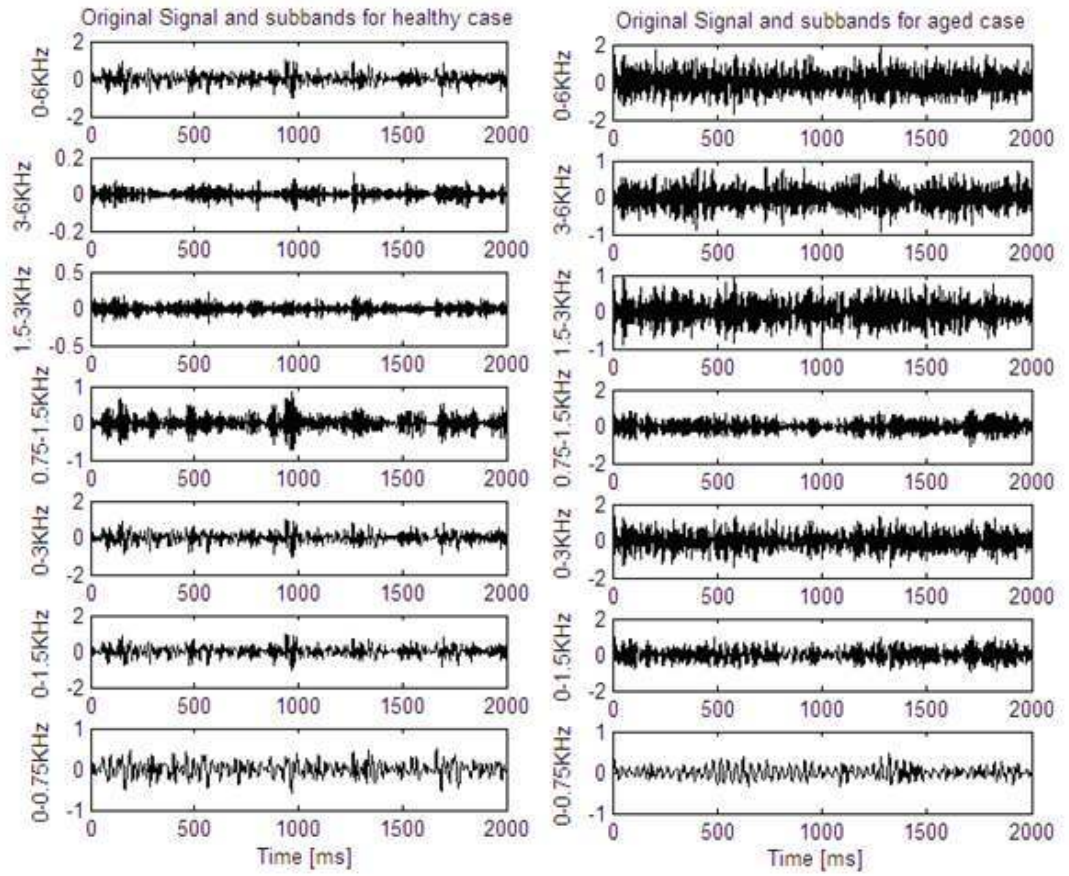

Figure 4.5 : The 5 level MRWA of 2 second samples of the vibration signal from cycles 0 and 7 at no load.

The 7 dimensional feature vectors are reduced using various dimensionality reduction techniques like principal component analysis, backward selection and single feature selection; and the success rate is checked for every dimension.

The HMM used in the experiment is a discrete 8 state left-to-right HMM as shown in Figure 4.6 where each state represents a cycle of the motor.

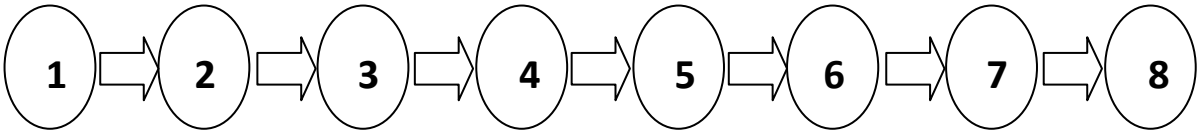

Figure 4.6 : 8 state left to right HMM

Since a discrete HMM is used, the observation output vectors $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ need to be discretized. In order to do the discretization, the training set feature vectors are clustered using

K-means clustering. During validation test, each feature vector is clustered into 20 clusters, i.e. the cluster it belongs to is found and the center of that cluster is taken as the emission(observation output). The whole experiment flow can be seen in Figure 4.4.

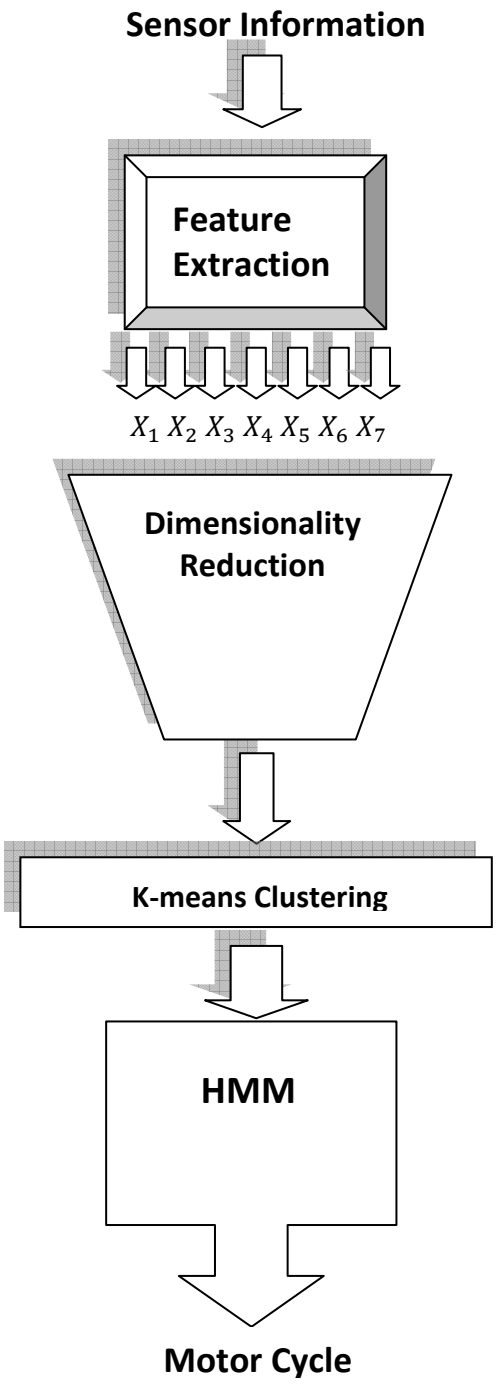

Figure 4.7 : The experiment flow

In this experiment, the vibration data need to be classified into 8 different aging states by the 8 state left-to-right HMM. The correctly classified percentage of the validation set is used as the measure of the classification. There are correlations between the 7 dimensional features. In order to find the best set of features to use for classification, dimensionality reduction techniques are used. In other words; the best classification will occur with the features that are directly related with the fault we

are interested in, thus dimensionality reduction techniques will be expected to keep the features that are related to certain frequency bands which characterize the fault, and remove the rest of the features.

Using PCA, the best performance of $95 \pm 1.9\%$ was obtained using only the first principal component (Table 8.1 , second column).

Using backward feature selection points out to the 5^{th} feature as the best feature with the classification accuracy of $99.64 \pm 0.2\%$ (Table 8.1 , third column). Backward feature selection eliminated features in the order of 6,7,4,3,1,2. It is notable that using all features results in classification accuracy of $94.5 \pm 1.3\%$, while using 7 PCA components achieves only $91.9 \pm 2.8\%$ accuracy.

Using a single feature at a time, best classification occurs ($99.6 \pm 0.2\%$) with the 5^{th} feature (Table 8.1 , fourth column.)

TABLE 4.1 : Percentage classification accuracy with different dimensionality reduction techniques.

i	First *ith* Principal components	Backward Feat.Select. *i* features	Only the *ith* feature
1	**95.02±1.9**	**99,64±0.2**	**91.03±4.8**
2	**64.1±2.9**	**98,6±1.0**	**95.5±1.3**
3	**68.4±1.5**	**98,3±1.2**	**93.2±0.7**
4	**72.7±3.8**	**97,94±1.6**	**92.5±2.1**
5	**81.3±3.0**	**97,61±3.0**	**99,64±0.2**
6	**86.8±4.5**	**94.8±0.3**	**92.07±1.3**
7	**91.9±2.8**	**97,5±1.3**	**91.6±1.6**

The fifth feature corresponds to the 1500-3000 Hz frequency band which is the band associated with the bearing damage [8, 41, 42]. Thus, the feature that represents the frequency band that has the highest correlation with the fault is detected.

4.2.2 Condition Monitoring

Given an observation sequence to an HMM, the Viterbi algorithm calculates the probabilities of the possible state sequences and the state sequence with the highest probability is chosen as the most probable state sequence, given the observation sequence.

Predictive maintenance method described in this study, greatly relies on this idea. The healthy condition is taught to HMM, and as the aging progresses by time, the increasing fault will create less probable observation by time; the increasing fault will create less probable observation output sequences thus decreasing the probability. A predetermined threshold value will separate the healthy condition from the condition that maintenance is necessary before a fault.

This application tries to foresee bearing damages and stator winding faults -the aging of the insulation of the stator winding- before the motor can be classified as a faulty motor.

Since this application is expected to have an applicability in industrial processes, relying solely on vibration data might cause robustness problems in some applications, thus the method is expected to work either with vibration or current (motor current) information.

The sensor information; current or vibration signal is preprocessed i.e., feature extraction using MRWA -5 level decomposition as seen in Figure 4.8 and Shannon Entropy to form the feature vectors. The future vectors are clustered and the corresponding cluster centers are passed to HMM as the observation sequence. The HMM's logarithmic probability outputs are observed to see if the values are below a predetermined threshold or not. The whole flow of the application is given at Figure 4.9.

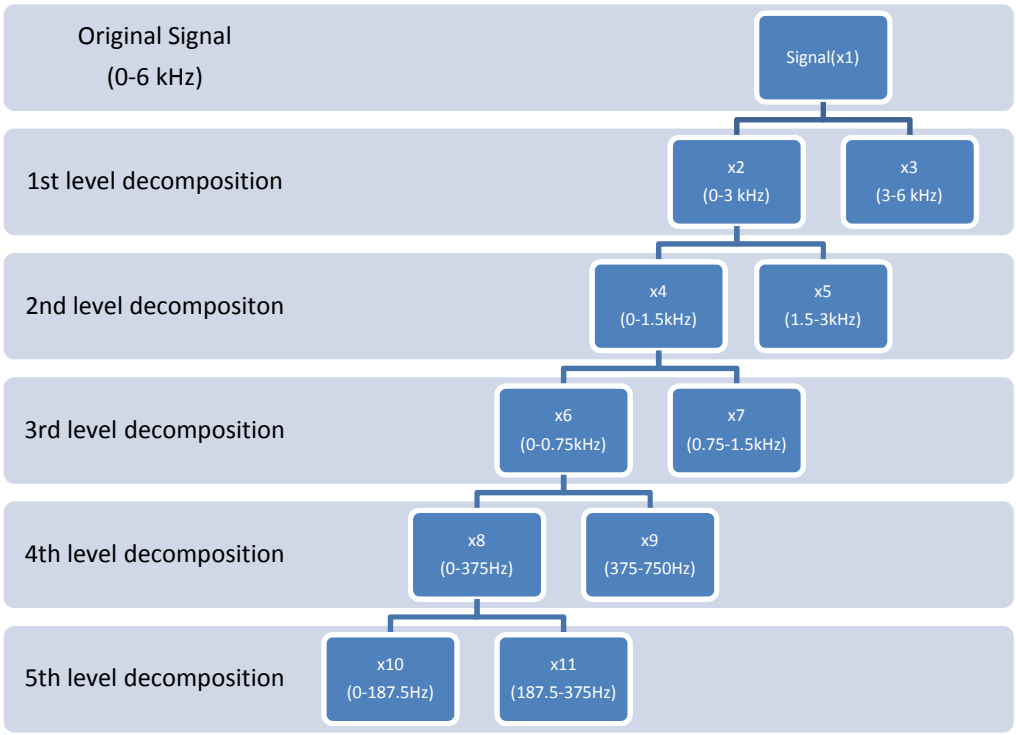

Figure 4.8 : 5 level MRWA decomposition of the signal forming the feature vectors

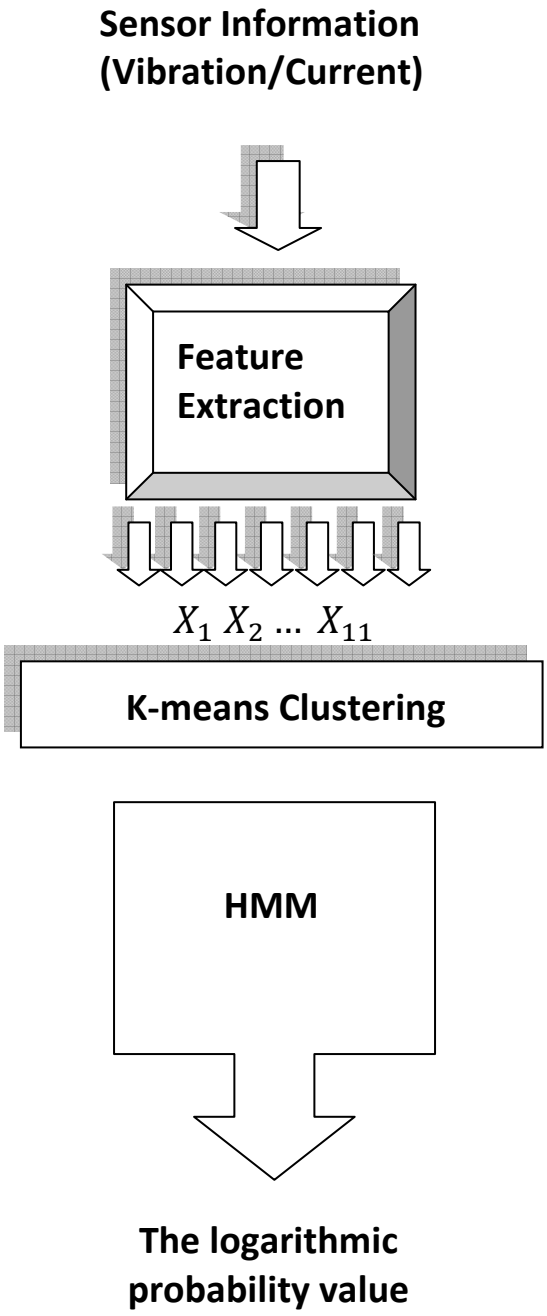

Figure 4.9 : The experiment flow

This process is repeated every 50 milliseconds. The 8 seconds of the cycle 0 data including each of the 5 different load conditions are given to the HMM as the training data and the rest of the data (2 seconds for each of the 5 different load conditions of cycle 0 and 10 seconds for each of the 5 different load conditions of the every other 7 cycles) is used for validation.

The HMM checks and updates the probability every 50 milliseconds. The other parameters of the system are given in Table 8.2.

Table 4.2 : The parameters of the system

Type of HMM	**4 state-left-to-right**
Number of Clusters	**65**
MRWA Level	**15**

Briefly, the signal is processed according to Figure 4.9 and the output is the probability of the data being in the healthy (cycle 0) condition. As the aging progresses, it is expected that HMM will yield smaller probability values.

The results of bearing damage - vibration of three motors and the result of bearing damage - current of a motor are given in Figure 4.10, Figure 4.11, Figure 4.12.

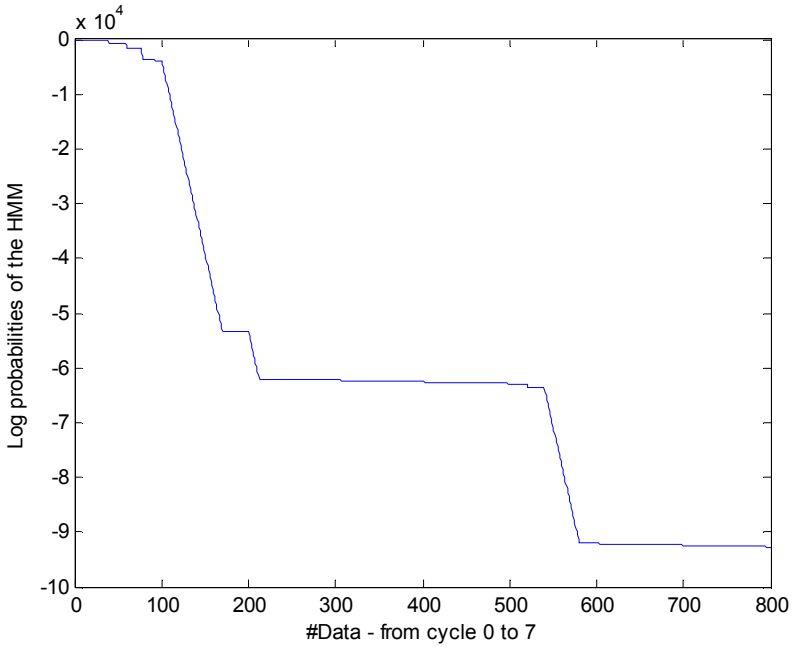

Figure 4.10 : Motor 11 vibration sensor for bearing damage

It is seen that in Figure 4.10, cycle 0 has the logarithmic probability of 0 (means p=1), and as the aging progresses further, the probability decreases. Between cycles 1 and 2 the probability value has a sharp drop down which can be accepted as the warning for the forthcoming fault [41,42]. Again another sharp drop down can be seen at the end of cycle 6, which is the starting point of the faulty operating which is not an acceptable operating condition [41,42].

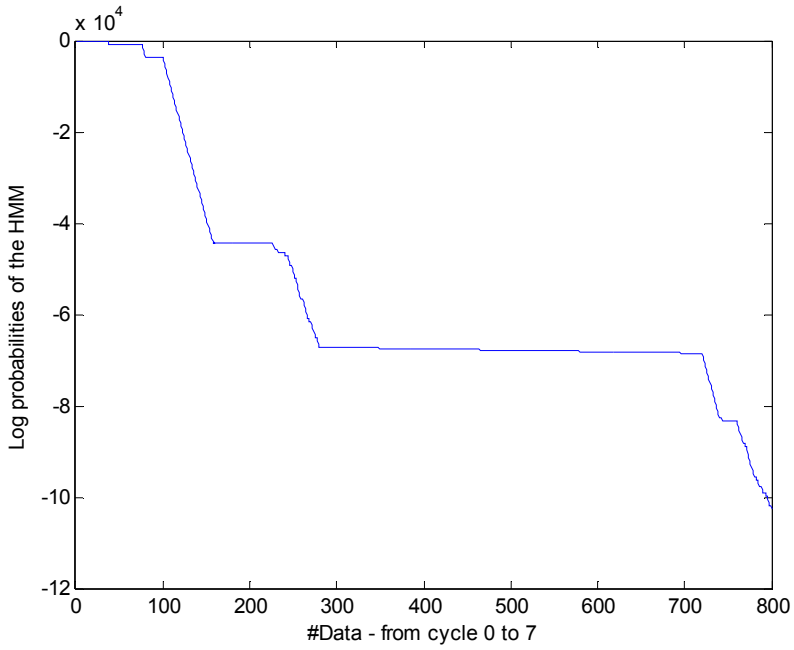

Figure 4.11 : Motor 12 vibration sensor for bearing damage

In Figure 4.11 the sharp drop downs can be seen at cycles 1,2 and 4. After the cycle 2, the motor starts building a considerable rotor fault and the working of cycle 7 is not acceptable for normal condition mode. Cycle 2 can be selected as the warning of the forthcoming fault.

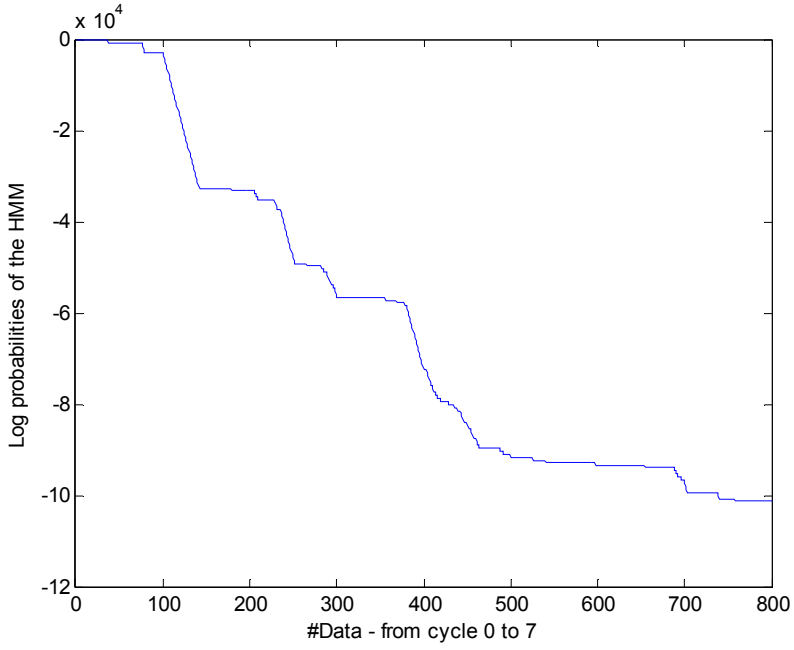

Figure 4.12 : Motor 13 vibration sensor for bearing damage

In Figure 4.12, after cycle 0, the decreasing probability trend shows the increasing severity of the fault. Cycle 4 can be selected as the warning of the forthcoming fault.

According to the above graphics, a global threshold value of -60000 can be selected which is acceptable for all of the three motors.

Figure 4.13 shows the HMM probabilities of the motor 11 according to the current sensor as the severity of the fault increases. Current sensor has a staircase like trend in terms of the decreasing probability values. Cycle 1 is the first appearance of the fault and it is the first drop down point in Figure 4.13. The probability drops down as the severity of the fault increases.

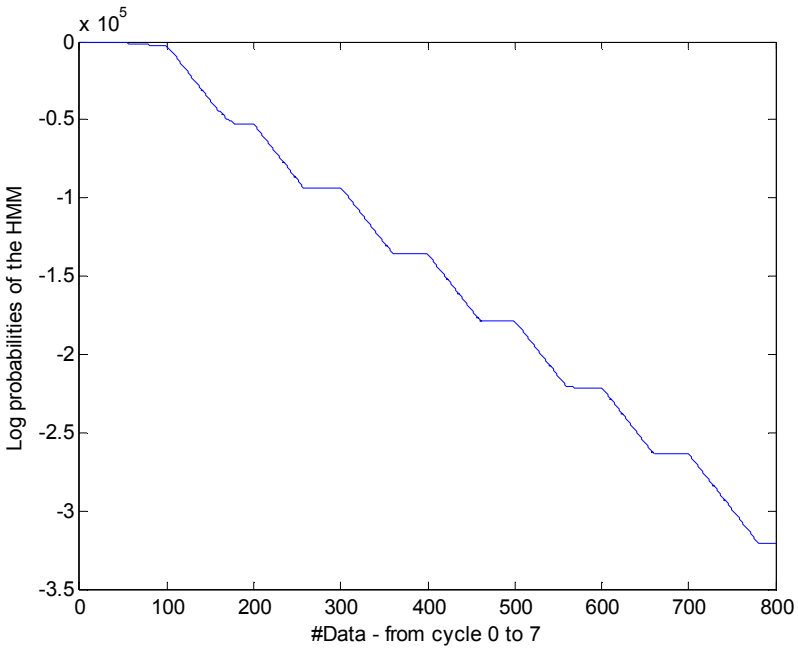

Figure 4.13 : Motor 11 current sensor for bearing damage

The results of stator fault-vibration , stator fault-current are given in Figure 4.14 and Figure 4.15 respectively.

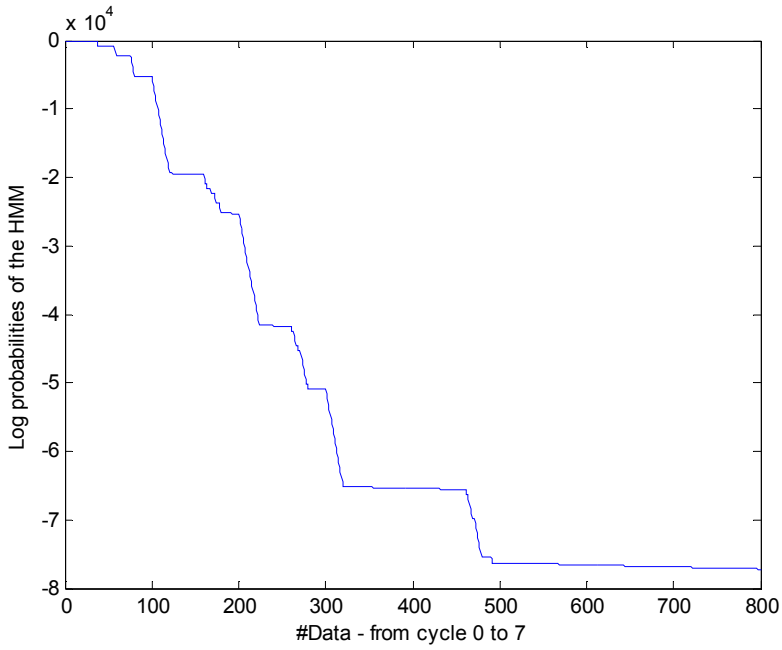

Figure 4.14 : Motor 1 Stator Fault – Vibration sensor

In Figure 4.14 it can be seen that the beginning of the fault is detected at cycle 1, and the increasing severity caused the decreasing probability trend. Cycle 4 can be taken as the last acceptable working mode and a warning should be triggered before that point.

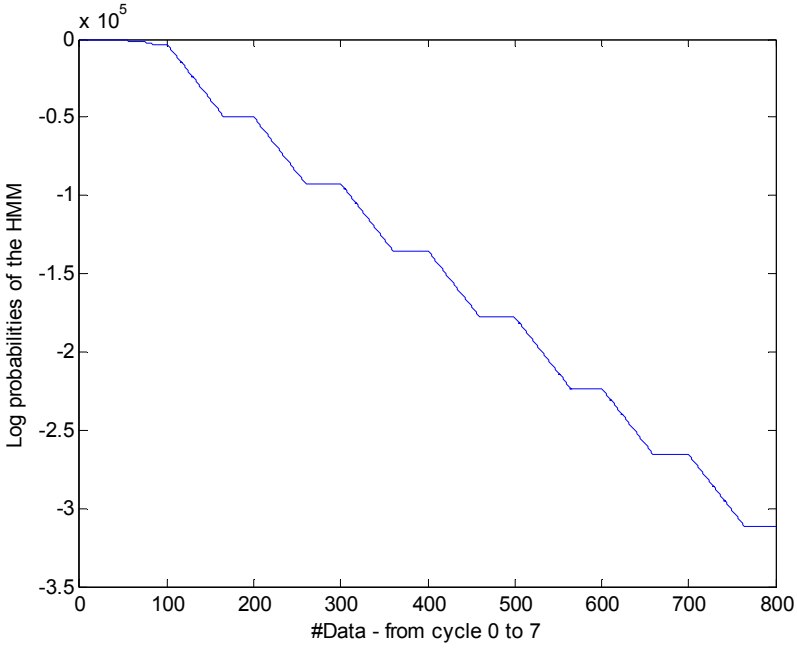

Figure 4.15 : Motor 1 Stator Fault – Current Sensor

Figure 4.15 shows the HMM probabilities dropping down with a staircase like function as the severity increases.

As the analysis of the above figures (Figure 4.10 to Figure 4.15) shows, HMM is a powerful tool for foreseeing the faults before they occur, which is a must of predictive maintenance.

5. CONCLUSIONS AND DISCUSSIONS

Two different fault related schemes were represented in this work.

In the first one, a new MRWA and HMM based fault characterization method was introduced. In order to test the method, rotor fault was chosen since the characteristics of a rotor fault is known [41,42]. The method used various dimensionality reduction techniques on the entropies of different levels(frequency bands) of the MRWA. The classification success was expected to be maximized when the features that are not strongly related with the fault are removed. The methodical removal of the features, namely the dimensionality reduction, resulted in a single feature(frequency band) with the maximum classification success.

This showed that the rotor bearing damage is directly related with 1.5-3 kHz frequency band in comparison to the rest of the applied MRWA bands, which is in concordance with the related literature [41,42].

The result of the application shows that this new method is valid and accurate for fault characterization.

As a future complementary work, this method can be expanded using wavelet packet decomposition (WPD) in substitution of MRWA. WPD provides a more detailed frequency band spectrum for analysis; thus more zoomed features, leading to more accurate results. The method can also be applied to different faults; stator faults, broken bar problems, misalignments etc.

In some of these faults, a simple wavelet analysis would be enough to visually discriminate the frequencies that directly represent the fault, unfortunately this is a rare case. Faults like the ones occurring in stator windings are not visually observable by using signal analysis methods. The difference of an observable and a non-observable fault character can be seen in Figure 5.1.

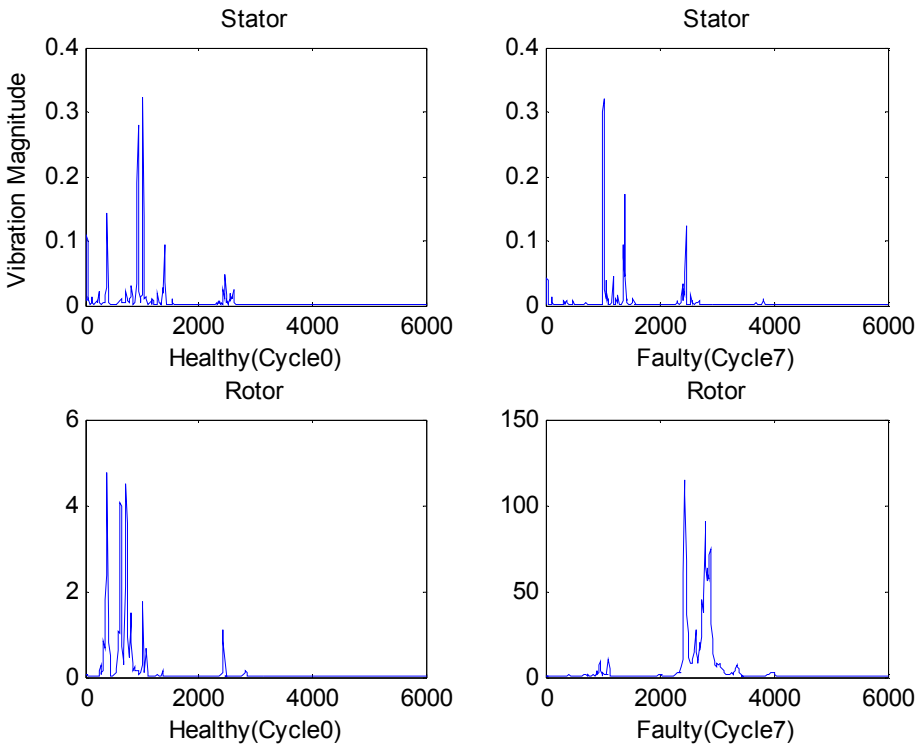

Figure 5.1 : The comparison of psd's of rotor and stator faults.

Figure 5.1 shows that the rotor fault character can easily be captured visually by the psd of the signal, however stator fault has no visually detectable anomaly.

Briefly, the benefit of this new method is the ability to find fault characteristics not only relying on signal analysis methods but also powering them up by tools of pattern recognition thus providing a new and accurate tool for fault characterization.

The second application of this study was a predictive maintenance method where even the slightest aging was detected way before a fault occurrence. The HMM was trained only with the healthy condition (cycle 0 – 2 seconds of the 10 seconds data for every load scheme). The motor was run with varying load during both training and validation.

Two different faults were tracked namely rotor (bearing) and stator faults. Also two different sensor types were used being vibration and current sensors.

Three motors' vibration data were present to test the rotor fault. As the aging progressed the logarithmic probability of the HMM (the logarithmic probability of the motor being in healthy state given the feature vector) diverged to large negative

values which means a 0 probability. The Figure 4.10, Figure 4.11 and Figure 4.12 shows the probabilities as the aging severity increases.

The first signs of the aging was detected with an obvious probability drop at the appropriate cycle.

Using current sensor, the bearing fault is much more obviously tracked as seen in Figure 4.13. Every cycle is represented with a stair drop in the HMM probability. Every step of the fault can be seen in the figure.

One can easily say; by comparing the Figure 4.10, Figure 4.11 and Figure 4.12 with Figure 4.13, that using current information for rotor fault detection is a better way than using vibration sensor. Vibration sensor is also more prone to be affected by environmental noises and inferences [43], especially when the system is an induction machine [43]. However current signal is more resistant to environmental noise [43]. Also using current sensors has an installation advantage on vibration signals, since current sensors are connected to cables outside the motor and in order to install the vibration sensors the machine needs to be disassembled and reassembled.

The stator (winding) fault was tracked using both vibration and current signals as seen in Figure 4.14 and Figure 4.15.

Two methods were successful since both showed every change in the developing fault. However current sensor was again better since it represented a staircase like trend, where each stair represented an aging cycle. The above explanations about the current and vibration sensors are also valid for stator fault tracking.

Moreover; the successful tracking of stator fault is an encouraging result in terms of the future work of the first application; the fault characterization. Since HMM is good at tracking the stator fault, which means HMM is good at handling the probabilities of stator fault features, then using dimensionality reduction techniques with HMM, the exact features(frequency bands) that have the strongest relation with the fault can be found. This means that the stator fault can be characterized in terms of frequency intervals.

The second application is a robust method for tracking the faults of induction motors. The system can give a warning that the aging of the motor has started, and increasing the warning level as the severity of the fault increases, which means that instead of costly and bulky periodic maintenance, predictive maintenance can be provided for

the induction motors. As a future work, the separate HMM models(rotor-stator faults) can be combined in a structure and various faults and also their combinations can be tracked and detected, which would be the universal condition monitor for induction motors with a very small learning time.

REFERENCES:

[1] **Thompson, S. P.**,2005. Michael Faraday: His Life and Work, Adamant Media Corporation.

[2] **Chapman, S. J.**, 1999. Electric Machinery Fundamentals, Mc Graw Hill.

[3] **Ocak, H., Loparo, K. A.**, 2005. HMM-Based Fault Detection and Diagnosis Scheme for Rolling Elements Bearings, *Journal of Vibration and Acoustics*, Asme.

[4] **Karatoprak, E., Şengüler, T., Ayaz, E., Çağlar, R., Şeker, S.**, 2007. Spectral and Statistical Based Modeling for Bearing Damage in Induction Motors, *The 6th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives* ,Cracow-Poland.

[5] **Şengüler, T., Karatoprak, E., Ayaz, E., Güllülü, S., Şeker, S.**, 2007. Entropy Approach Using PCA for Sequential Accelerated Aging Processes in Electrical Motors, *The 6th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, Cracow-Poland.

[6] **Şeker, S., Karatoprak, E., Kayran, A. H.**, 2007. Stationary wavelet transform for fault detection in rotating machinery, *SPIE Optics East 2007*, Seaport World Trade Center, Boston, MA,USA .

[7] **Karatoprak, E., Şengüler, T., Ayaz, E., Şeker, S.**, 2007. Comparisons of the continuous and Discrete Wavelet Transforms for Potential Failure Detection in Electric Motors, *ELECO 5th International Conference on Electrical and Electronics Engineering*, Bursa, TURKEY

[8] **Karatoprak, E., Şeker, S., Çataltepe, Z., Şengüler, T.**, 2007. Using Bayes Decision Rule for Motor Fault Detection, *SIU (IEEE 15th Signal Processing and Communication Applications Conference)*, Eskisehir, Turkey.

[9] **Karatoprak E., Şengüler, T., Şeker, S.**, 2006. PMDC Motor modelling, control , fault detection and classification based on artificial neural networks, *Turkish National Automatic Control Committee 2006 Meeting* Ankara.

[10] **Rabiner, L. R.**, 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Fellow, IEEE, *Proceedings of the IEEE*, **vol 77**,No.2.

[11] **Alpaydın, E.**, 2004. Introduction to Machine Learning, The MIT Press Cambridge, Massachusetts, London, England.

[12] **Jurafsky, D., Martin, J. H.**, Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition.

[13] **Viterbi, A.**, 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Transactions on Information Theory*, **Volume 13**, Issue 2, Page(s):260 - 269

[14] **Forney, G.D. Jr.**, 1973. The viterbi algorithm, *Proceedings of the IEEE* **Volume 61**, Issue 3, Page(s):268 – 278

[15] **Baum, L. E., Petrie, T., Soules, G., Weiss, N.**, 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Statist*, **vol. 41**, no. 1, pp. 164—171.

[16] **Dempster, A. P., Laird, N. M., Rubin, D. B., Roy, J.**, 1977. Maximum likelihood from incomplete data via the EM algorithm.", *Stat. Soc.*, **vol. 39**, no.1, pp. 1-38.

[17] **Ocak, H. Loparo, K. A.**, 2005. HMM-Based Fault Detection and Diagnosis Scheme for Rolling Element Bearings, *Journal of Vibration and Acoustics*, **Vol.127**/305

[18] **Ghahramani, Z.**, 2001. An introduction to Hidden Markov Models and Bayesian Networks, *International Journal of Pattern Recognition and Artificial Intelligence*, **15**:9-42

[19] **The New York Times**, 19 October 1931

[20] **Bi, J., Bennett, K. P., Embrechts, M., Breneman, C. M., Song, M.**, 2003. Dimensionality Reduction via Sparse Support Vector Machines, *Journal of Machine Learning Research* **3**, 1229-1243

[21] **Globerson, A., Tishby, N.**, 2003. Sufficient Dimensionality Reduction, *Journal of Machine Learning Research* **3**, 1207-1331

[22] **Shlens, J.**, 2005. A tutorial on Principal Component Analysis, Systems Neurobiology Laboratory, Salk Institute for Biological studies and Institute for Nonlinear Science, University of California,San Diego.

[23] **Smith, L. I.**, 2006. A tutorial on Principal Component Analysis, University of Otago, New Zealand.

[24] **Duda, R. O., Hart, P. E., Stork, D. G.**, 2000. Pattern Classification, 2nd Edition Wiley.

[25] **Bishop, C. M.**, 1995. Neural Networks for Pattern Recognition, Oxford University Press ; Oxford England.

[26] **Cover, T. M., Thomas, J. A.**, 1991. Elements of Information Theory, New York, Wiley.

[27] **Shannon, C. E.**, 1948. A Mathematical Theory of Communication, *The Bell System Technical Journal*, **vol:27**,p379-423.

[28] **Schneier, B.**, 1996. Applied Cryptography, Second Edition, p234 John Wiley and Sons.

[29] **Oppenheim, A. V., Schafer, R. W.**, 1975. Digital Signal Processing, Prentice Hall 1975.

[30] **Boyce, W. E., DiPrima, R. C.**, 2005. Elementary Differential Equations and Boundary Value Problems, John Wiley&Sons, New Jersey.

[31] **Freeman, A.**, 2003. The Analytical Theory of Heat, Joseph Fourier, (published 1822, translated 1878, re-released 2003), Dover Publications 2003 unabridged republication of the 1878 English translation by Alexander Freeman of Fourier's work Théorie analytique de la Chaleur 1822.

[32] **Reed, D. G.**, 2003. The Arrl Handbook for Radio Amateurs, American Radio Relay League Cp.18.

[33] **Daubechies, I.**, 1992. Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics.

[34] **Mallat, S.**, 1999.A Wavelet Tour of Signal Processing, Academic Press, 2nd Edition.

[35] **Vidakovic, B., Mueller, P.**,2007. Wavelets for Kids,Duke University.

[36] **Akansu, A., Haddad, R.**, 2001. Multiresolution Signal Decomposition: Transforms, Subbands and Wavelets, Academic Press, Boston MA.

[37] **Ayaz, E., Öztürk, A., Şeker, S.**, 2006. *Continuous Wavelet Transform for Bearing Damage Detection in Electric Motors, IEEE* Melecon Benalmadena-Malaga, Spain.

[38] **Ho, K. H., Newman, S. T.**, 2003. State of the art electrical discharge machining (EDM), *International Journal of Machine Tools and Manufacture*, **Volume 43**, Number 13,pp.1287-1300 (14) Elsevier.

[39] **Şeker, S., Ayaz, E., Upadhyaya, B. R., Erbay, A. S.**, 2000. *Analysis of motor current and vibration signals for detecting bearing damage in electric motors*, MARCON, Maintenance and Reliability Conference, Knoxville.

[40] **ANSI C50.32-1976 , IEEE Std 117-1974**, IEEE Standard Test Procedure for Evaluation of Systems of Insulating Materials for Random-Wound AC Electric Machinery.

[41] **Bonnet, A. H., Root**, 2000. Cause AC Motor Failure Analysis with a Focus on Shaft Failures, *IEEE Transactions on Industry Applications* **36**(5),1435-1448.

[42] **Şeker, S., Ayaz, E.**, 2002. A study on condition monitoring for induction motors under the accelerated aging processes, *IEEE Power Engineering Review* **22**(7), 35-37.

[43] **Leea, S. K., Whitea, P. R.**, 2002. The Enhancement of Impulsive Noise and Vibration Signals for Fault Detection in Rotating and Reciprocating Machinery, Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, England.

[Url-1] <http://www.wolfram.com/webresources.html> , accessed at 22.11.2007.

VITA

Candidate's full name: **Hamid Erinç KARATOPRAK**

Place and date of birth: **Ispir 18/09/1982**

Permanent Address: **Istanbul Technical University, Faculty of Electrical&Electronical Engineering, Department of Electrical Engineering, MASLAK,ISTANBUL**

Universities and Colleges attended: **Istanbul Technical University**

Publications:

- **Karatoprak E**., Şengüler T., Ayaz E., Çağlar R., Şeker S., 2007. Spectral and Statistical Based Modeling for Bearing Damage in Induction Motors, The 6th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, Cracow-Poland.
- Şengüler T., **Karatoprak E**., Ayaz E., Güllülü S., Şeker S., 2007. Entropy Approach Using PCA for Sequential Accelerated Aging Processes in Electrical Motors, The 6th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, Cracow-Poland.
- Şeker S., **Karatoprak E**., Kayran A. H., Şengüler T., 2007. Stationary wavelet transform for fault detection in rotating machinery, SPIE Optics East 2007, Seaport World Trade Center, Boston, MA, USA.
- **Karatoprak E**., Şengüler T., Ayaz E., Şeker S., 2007. Comparisons of the continuous and Discrete Wavelet Transforms for Potential Failure Detection in Electric Motors, ELECO 5th International Conference on Electrical and Electronics Engineering, Bursa, TURKEY
- **Karatoprak E**., Şeker S., Çataltepe Z., Şengüler T., 2007. Using Bayes Decision Rule for Motor Fault Detection, SIU (IEEE 15th Signal Processing and Communication Applications Conference), Eskisehir, Turkey.
- **Karatoprak E**., Şengüler T., Şeker S., 2006. PMDC Motor Modeling, Control, Fault detection and Classification Based on Artificial Neural Networks, Turkish National Automatic Control Committee 2006 Meeting Ankara.